

STIC-Biotech/ChemLib

6/1/44

From: Slobodyansky, Elizabeth
Sent: Tuesday, February 26, 2002 5:59 PM
To: STIC-Biotech/ChemLib
Subject: 09/724,126

Please search for case 09/724,126:

SEQ ID NOs: 1 and 2 against commercial ~~and interference~~ databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652
CM1 10D11
703-306-3222

mail box 10C01

RECEIVED
FEB 27 2002
STIC

if Contact:
Sheppard

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 3/2/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:40:50 ; Search time 95.27 Seconds
(without alignments)
14995.520 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattccgacgagg.....aatttgattgtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2.6/ptodata/2/ina/PCRUS_COMB.seq.*
- 6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4435.4	70.3	6395	US-08-982-956-1	Sequence 1, Appli
2	4435.4	70.3	6395	US-09-228-317-1	Sequence 1, Appli
3	996.2	15.8	1001	US-08-982-956-2	Sequence 2, Appli
4	996.2	15.8	1001	US-09-228-317-2	Sequence 2, Appli
5	57.4	0.9	7218	US-08-232-463-14	Sequence 14, Appli
6	52.2	0.8	7218	US-08-232-463-14	Sequence 14, Appli
7	44.4	0.7	7333	US-09-138-024-21	Sequence 21, Appli
8	44.2	0.7	72928	US-09-009-913-1	Sequence 1, Appli
9	42.2	0.7	325	US-08-991-789A-236	Sequence 236, App
10	41	0.6	1517	US-08-963-602-6	Sequence 6, Appli
11	40.4	0.6	152331	US-09-128-155-16	Sequence 16, Appli
12	39.8	0.6	2861	US-08-299-953-1	Sequence 1, Appli
13	39.8	0.6	2861	US-08-459-415-1	Sequence 1, Appli
14	39.8	0.6	2861	PCT-US95-11231-1	Sequence 1, Appli
15	39.8	0.6	3881	US-08-299-953-2	Sequence 2, Appli
16	39.8	0.6	3881	US-08-459-415-2	Sequence 2, Appli
17	39.8	0.6	3881	PCT-US95-11231-2	Sequence 2, Appli
18	39.8	0.6	4376	US-08-119-125A-1	Sequence 1, Appli
19	38.8	0.6	246240	US-08-724-394A-20	Sequence 20, Appli
20	38.8	0.6	246240	US-08-724-394A-21	Sequence 21, Appli
21	38.8	0.6	246240	US-08-724-394A-22	Sequence 22, Appli
22	38.4	0.6	5703	US-09-280-590A-36	Sequence 36, Appli
23	38.4	0.6	18596	US-09-318-448-11	Sequence 11, Appli
24	37.8	0.6	19011	US-08-310-356-36	Sequence 36, Appli
25	37.8	0.6	19557	PCT-US92-06300-1	Sequence 1, Appli
26	37.6	0.6	72928	US-09-009-913-1	Sequence 1, Appli
27	37.4	0.6	1333	US-08-543-246B-15	Sequence 15, Appli

C 28	37.4	0.6	1387	4	US-08-543-246B-1	Sequence 1, Appli
C 29	37	0.6	1113	1	US-08-341-538A-1	Sequence 1, Appli
C 30	37	0.6	1113	2	US-08-725-518-1	Sequence 1, Appli
C 31	37	0.6	1618	3	US-08-889-108-1	Sequence 1, Appli
C 32	37	0.6	1618	3	US-08-889-108-3	Sequence 3, Appli
C 33	37	0.6	1618	4	US-08-120-601B-1	Sequence 1, Appli
C 34	37	0.6	1618	4	US-08-120-601B-3	Sequence 3, Appli
C 35	37	0.6	1618	5	PCT-US94-10358-1	Sequence 1, Appli
C 36	37	0.6	1618	5	PCT-US94-10358-3	Sequence 3, Appli
C 37	37	0.6	4875	1	US-08-460-739-1	Sequence 1, Appli
C 38	36.6	0.6	1480	4	US-09-290-640-65	Sequence 65, Appli
C 39	36.6	0.6	3784	1	US-07-623-033-1	Sequence 1, Appli
C 40	36.4	0.6	1222	4	US-08-543-246B-5	Sequence 5, Appli
C 41	36.2	0.6	9636	1	US-08-323-170B-1	Sequence 1, Appli
C 42	36.2	0.6	19124	2	US-08-487-826B-13	Sequence 13, Appli
C 43	35.4	0.6	252	2	US-08-332-766A-28	Sequence 28, Appli
C 44	35.4	0.6	2135	4	US-08-430-286A-1	Sequence 1, Appli
C 45	35.2	0.6	289	4	US-09-007-005-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-08-982-956-1
; Sequence 1, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
; US-08-982-956-1

Query Match 70.3%; Score 4435.4; DB 2; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

Oy 590 ttcaaggccgtcgtaaaagtctgcctctgtctc-tccgaccggccacaggttttcgct 648

Qy 2809 ataagttcttctgacttcagaggtatgaacttgcgaggtcttttaacaagacca 2868
 Db 2228 ACAAGTTCTTGTACTGTACTTCAGAGATATGAACCTTACTGCTTTTAAACAAGACCA 2287
 Qy 2859 tatctacaaaagaccagagatttgattaaacaataataacactaataagaataagcttc 2928
 Db 2288 TATCCACAAAAGACCAGATTGTGATTAAACAGATATAATACATTAATAGAAAGAAATGCTTC 2347
 Qy 2929 aggtcctcatctattgtggtagcgttatgtacctggagtggaatgtgaccaag 2988
 Db 2348 AGGTCTCATCTATTGTGGAGAACCTTATGTACCTGGAGTGGAAATGTTACACAG 2407
 Qy 2989 aagaggtcacatgagagaaatcattccttgccttgattgaacccatgccacacagtg 3048
 Db 2408 AGGAGTTTAAATGAGAGAGATTACTCACTTACTTTGCTTGGAGCCCATGCCACAGTG 2467
 Qy 3049 ccattgccaataattacctgagaatgaaaataatgaactggcttagaatagtcataa 3108
 Db 2468 CCATGCCAGAAACCTACTCTGAGAAGCAAAATATGAACCTGGCTTAGAATGTCTATAA 2527
 Qy 3109 acaagtgccacatttaagaaccaggtgatcagggccatggagtttatgaactaaaag 3168
 Db 2528 ACAAGTGGCCACATTTAAGAAACCAGGTGTCTGGGCCATGGAGTTTATGAATTTGAAG 2587
 Qy 3169 atgaactactgaagacttcaatgtacttttattacttactccaaacacccagcatagca 3228
 Db 2588 ATGAATCACTGAAGACTTCAATATGTACTTTTACCATTATTTCTAAAACACAGCATAGCA 2647
 Qy 3229 aggtggaacatatgcagaagaaaggagaaacaaagaaacaaagatgaagcattgcgc 3288
 Db 2648 AGGTGAAACATATGCAGAAGAAAGAGAAACAAAGAAATAAAGATGAAGCATTTGCCGC 2707
 Qy 3289 caccacacccctcctgaattctgcctgtcttcagacaaagtattaaacctctcaactgtg 3348
 Db 2708 CGCCACCTTCCTCCAGAGTCTGCCTGTCTTCAGCAAAAGTAGTCAACCTGTCTCAGCTGTG 2767
 Qy 3349 atactgatgatcaattctcagaccgttattgagcggcgaataacacacagattcttaact 3408
 Db 2768 ATGTATGATATACATCTCAGGACCATCTTTGAGCGGGCAGTGACACGGAGTCTATATC 2827
 Qy 3409 tgtgaccgaaggatgctccaaatggctttctattcttgcattggcattgggtttactagaag 3468
 Db 2828 TGTGGACAGAAGGATGTGCAGATGGCTTCCATATATTGGCACTGGCTTGGTGAAG 2887
 Qy 3469 agaagcaacacttcaaaaagcctcgtgaagaaagttaacattgaactttatcataagg 3528
 Db 2888 AGAAGCAGCAGCTTTCAGAAAGCTCCTGAAGAGGAGTGGCTTTTGACTTTTACCATAAAG 2947
 Qy 3529 ctccaagattggagttcagccat- - - - - gaatatcaaatgcttttggaaaaac 3579
 Db 2948 CTTCAGATTGGGAAGTTTACGCCATGAATGTCAGATATATCAAAATGCTTTGGAAAGAC 3007
 Qy 3580 tcaagggaattccccagttagaagccagagagacatgaacgtggatacttcagatgt 3639
 Db 3008 TCAAAAGGAATCCCCAATTAGAAGCCAGAGGACATGATACATGGATACTCCAGATGT 3067
 Qy 3640 ttgacacagtgaaagcattagagaaaaatctgtttaaattgtagcaaccacatcagat 3699
 Db 3068 TTGACACAGTGAAGCGATTAAAGAAAAATCTTGTGTTAGTTGTGGCAACCACTTCAGGAC 3127
 Qy 3700 cggaaattcaagaatgagattactcatgatataaagaaaaacagaaagaaaaagaa 3759
 Db 3128 TGGAGTGATTAAAGATGAGGAGATTACTCATGTATTAAGAAAGGCAGAAACGGAAGAAA 3187
 Qy 3760 aagctgaagctgctaggctacatgcgcgaagatcatggctcagatgctgccttacaga 3819
 Db 3188 AAGCTGAGGCCCTTAGGCTTCTGCCAAGAGATCATGCCCCAGATGCTGCTTTACAGA 3247
 Qy 3820 aaaaacttaattgaactcataactcatctgtatgacaaatcatatcagaaaatgacctgaaag 3879
 Db 3248 AAAACTTCAATTGAACCCACAAACTCACTGTATGATTAATACGTACAGAAAGTAACAGGGAAG 3307

Qy 3880 aagattccattatggagggaagagagaccacccagcagtcagtcactactctagaaattgctt 3939
 Db 3308 AAGACTCCATTATGGAGGAAGAGAGACACCTTCAGCAGTCACTAGAGCTCTTAGAATTTGCTC 3367
 Qy 3940 tgggtctcaaacgggtccatctgttactgaaagagggtgctgagctgcatcctcttgcc 3999
 Db 3368 TGGGCCCTTAACGGGCCCGGCTGTACCGAAAAGGAGGTGCTACGTCGATCCTCTTGCC 3427
 Qy 4000 aagaagaacagaggtgaaaaatagaaaaatagccatggttatctcagcctgtgtccaga 4059
 Db 3428 AAGAAGAACAAGAGGTAAANAATAAGANAATAATGCCATGGTATTGTACAGATGTGTGCAGA 3487
 Qy 4060 aatctactgctttaaaccacagcaggggaaacoccatagaaactctcaggagaagccctag 4119
 Db 3488 AATCCACGCCCTAACCCAGCACAGAGGGAAGCCTGTGGACCATTAGGGGAACACACTGG 3547
 Qy 4120 accacttttcatcgatccagacttggtcatataggaaacttatcagaagagtgtggtcatg 4179
 Db 3548 ACCCTCTTTTCATGATGCCAGACTTGGCATGGAACCTTATACAGGAAGCTGTGCTCATG 3607
 Qy 4180 taatgcacagctgtctggcagaagtatttgaagctgtacagctgagctctcagcagc 4239
 Db 3608 TAATGTCACAGTGTCTGGCAGAAATTTTGAAGCTGTGCAGCTGAGCTCGCAGCAGC 3667
 Qy 4240 gattcatgttgaccttttgccttgaaagtggaataatctttgacctctttgcaaat 4299
 Db 3668 GCATTCACTGATAGCTGTTTGACTGAGAGCGCGGAGTACCTATGCGCGCTCTGCAAGT 3727
 Qy 4300 ctctgtgcaataacttgatccccattattcttgcacacctcaaaaagataaaacagtgaga 4359
 Db 3728 CTCTCTGCAACACTGTATCCCATCATCCTTTTGCAGCGCAGAGATCAACAGTGA 3787
 Qy 4360 atgcagatgctctgtcctaaactttgacctggcagctgtgagatacacagctgttctgcca 4419
 Db 3788 ATGCGGAGGCTCTGTGCTCAACTTTTGACCTTGGCCCGGTGGATACAGACTGTCTTGCCA 3847
 Qy 4420 gaatatcaggttataataaagacatgctaaagagaaacac- - - caattcctatttct 4476
 Db 3848 GAATATCGGTTATATATATAAGCATGCTAAAGGAGAAAGCCCGCAGAGTTCCTCTCTGT 3907
 Qy 4477 ttaatacagaagatggagattcttactttggagttccattccactcctgagttttgggttg 4536
 Db 3908 TTAATCAAGGAATGGGGATTCAACTTTTGAGTTTTCATTCATCCTCAGTTTTGGAGTTC 3967
 Qy 4537 agtcttcgattaaatcaaatagcatacaagaaatggttattctcttcttgcacaacaa 4596
 Db 3968 AGTCTTTCGGTGAATATTCAAATAGTATCAAGGAAATGGTATTCTCTTCGCCAACAA 4027
 Qy 4597 ttatatgaattggattgaagtgccacctgatgaagggaatccctcagctcccatgctga 4656
 Db 4028 TTTACAGAAATGGCTTGAAAGTGCCCTCCTGATGAACTAGACCCACAGAGTGCCCATGATGA 4087
 Qy 4657 cctggagcactgctgttccactatccaggcaattgaaaatctatttggagagatgaagaa 4716
 Db 4088 CCTGGAGCAGCTGTGCTTCCATCCATCCAGCAATCGAAAAACCTGTTGGAGATGAAGAA 4147
 Qy 4717 aacctgtgttgagcaacttcaaaaataggcagcataatggtctgaaagcatttaagcagt 4776
 Db 4148 AACCTTATTTGGAGCACTTCAAAATAGACAGCATGCGGTCTCGAAGCGCTGATGCACT 4207
 Qy 4777 ttgcagttgcagagagattacctgctcctcaggttcctgatacagaacacatctgttctgc 4836
 Db 4208 TTGCAGTTGCACAGAGGGCTACCTGCGCTCAGGTCTTGTATACACAAACATCTGCTCGGC 4267
 Qy 4837 ttctatcagttgttcttccctaaataaaaaatcagaagatacaccatgctctctctatag 4896
 Db 4268 TCCTGTGAGTTATTCTTCCCTAACCTGCAATCAGAAAAATACACAGGCTCTTGTCTGTGG 4327
 Qy 4897 atctgttctaatgttttgggtggtgctgtgtagcattcccaatcctctgtattggatgacc 4956
 Db 4328 ATCTCTTCCATGTCTGTGCGGCGAGTCTTAGCGTTTCCCATCTCTGTATTATTTGGATGACA 4387
 Qy 4957 ctgtgtatctgcagccttcttcagttagtttcttctataaccacaccttctcttctcatt 5016

Db 4388 CCGTGGATCTGCAGCGCTGCCACTTAGTTCTTCATATAACCACTCTATCTCTTCCATC 4447
Qy 5017 tgataccatggaacacatgcttcagatactacttaacagtagaca-----cag 5064
Db 4448 TGATCACCATGGCGCATGCTTTCAGATACCTCTTACAACAGATACAGATGCTGCTCCAG 4507
Qy 5065 gctaccccttgcaggttcaagaagacagtagaagagcctcattccgcattcttcttct 5124
Db 4508 GGCGCGCTGTCTGAGGGTGAAGAGATAGTAGGAGGCTCGTGTGCATCTGCTTTCT 4567
Qy 5125 ttgcagaaattctcaataacacagtggtccattggtggtgatattcctggctgggtatt 5184
Db 4568 TTGTGGAAGTGTGCGACACAGACGGCTCACTAGTGGTGGGTGCTCCGGCTGGTACC 4627
Qy 5185 tgggtgctcactgaagaatggtcatccccccttatcttcgctgctgctattgttttcc 5244
Db 4628 TGTGGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCCGCTGTGCTGCACCTGCTTTTC 4687
Qy 5245 actatttacttgggttaactccgctgaggaactgcataccaaattctgcagaagagagt 5304
Db 4688 ACTATTACTTGGAGTAGCTCCGCCCTGAAGAACTGTTGGCAATTCGTGGAAGGAAAT 4747
Qy 5305 acagtgcactctgactatctatctttaccatacaaaattgttccctctctccaggaa 5364
Db 4748 TCAGTGCACCTCTAGCTATCTATCTTACCACAAATTTGTTCTCTGCTTTCCAGGAAT 4807
Qy 5365 attggatactgttaagcccttctccagaggtggtggtgcagatcctgccttaaaact 5424
Db 4808 ATTGGGATACCATTAAGCCCTTACTACAGAGGTGGTGGAGATCCCTGCTTACTCAAGT 4867
Qy 5425 gttgaagcaaaaaacacgctggtcaggtacccctgagaaagaaatagttgataagc 5484
Db 4868 CTTTGAAGCAGAAAGTGTGTGGTGTGAGTACCTTGAAGAAAGAAATAGTTGATAGAG 4927
Qy 5485 ttctgatactatagctgctcctgaatcaagctctctcattcaggtgccacggtctg 5544
Db 4928 TTCTGAGGACTACAGCTGTCTTCTAAATCAGGCTTCTCACTTGGTGTCCACGGTCTG 4987
Qy 5545 cagatgatgagcaaaagcctctgctcctctctctctctctctctctctctctctctctc 5604
Db 4988 CAGATGATGAGCAAAAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 5047
Qy 5605 agaaacttgcctgcaggaaaattgtgaacggggaagaggtgagcttgcatctttcaag 5664
Db 5048 AGAACATCTCTTGGCCAAGAAATAGTAAAGTGGGAAGAGGTTGAGCGCTGCGTTTTCATG 5107
Qy 5665 cactcactgtgagcgcaggtctgatttctcctaaatacagaaatccgaggtgctc 5724
Db 5108 CGTTTCATTTGTGTGGAGTCTGCATTTTCTTAAATCCAGAAATCCAGAAATCGAGGTGCTCC 5167
Qy 5725 tgggtgaaggttaaaagcagaggtgctcctctctcctcctcctcctcctcctcctcctc 5784
Db 5168 TGGTGAAGGAAAAGCCAGAGGCTGTGCCTTACCCAGCCCTTACTTGGATGAATATGAG 5227
Qy 5785 aaacagacctggcctgaagaggggcaacccccctcatttctcgtgagcgggtatcgga 5844
Db 5228 AAACAGCCAGCGGCTTAAAGAGAGAAACCCACTTCAATTTATCTCGGAGCGGTATCGGA 5287
Qy 5845 agctcatttgcctggaacacactgcattatagaagagattgctagagcgaagaga 5904
Db 5288 AGCTGCATTTGGTGTGGCAACACATGCTATTAAGAGAGATTGCTCGGAGCCAGGAGA 5347
Qy 5905 ctaatcagatgttatttgattcaactggcagttactgtgagctccaaactcctcctcaag 5964
Db 5348 CTAATCAGATGCTATTTGGATTTAACTGGCAGTTACTCTGAGGCTTCTGCTCCCTCAAG 5407
Qy 5965 acaatcacaaatgacacagtagtaaggtgattcaaaattatggaatactttctgag 6024
Db 5408 ACAATCATGAGTCACATCAATAAAGACTGATCTAAAATTTCTAGAGAACTTTCTGAGG 5467
Qy 6025 gctggaaagtattgaggggtcttttgcctcatgtccaggttcaactacatacaataa 6084

Db 5468 ACGGGGAAGTATTGGAGGTCTTTTGATCCATGTCCAGATTCCACACATTAATAAAAT 5527
Qy 6085 atttcttaagtgaattgtcttcaattagcaaacatagtcttcacagggaaaa-aggac 6143
Db 5528 ATTCTTTATGAATATTGCTTTCAATTATCAAAACATAAGCTTCAAGGGAAAAACAAGAC 5587
Qy 6144 atagatcaatcgt-----ttt 6160
Db 5588 ATAGATTAAATGTTTATGTTTAGAACACATAAAGAAATGCTTGTTCATCCAAAGTGTCTAT 5647
Qy 6161 atgtctgtagtattccaggaatttattccctctcaaatgtctcatttctatttt 6220
Db 5648 TTCTGCTAATATTCCAGAAACTCTTTCCCTTCATAACTGCTTAGTTTCATTTTCATAT 5707
Qy 6221 catccactggtagatgaagtcacgtcaaacagttgtagacattttatgtgttggttaac 6280
Db 5708 CACCACCTGGTAAATAGGTCACATTAAGCATTTGTGGACATTTCTCCATCTGGGTAAAC 5767
Qy 6281 tctctgcaatttctatttgggtttt 6307
Db 5768 ATCTCTGCACCTTTGTATTGTTGGTGT 5794

RESULT 2
US-09-228-317-1
; Sequence 1, Application US/09228317
; Patent No. 6159732
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P. C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/228,317
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
US-09-228-317-1

Query Match 70.3%; Score 4435.4; DB 3; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

Qy 590 ttcaggggccgctgtaaaagtgtcgtccctgtctc-tccgaccgcccacaggtttccgct 648

Db 4388 CGGTGATCTGCAGCGCTGCGACCTTACTTCTCATATAAACACACCTCTATCTCTCCATC 4447
Qy 5017 tgatcaccatgacacatgcttcagatctactactactacagtagaca-----cag 5064
Db 4448 TGATCATCATGGCGCACATGCTTCAGATACTCTCTTACACACATACAGATCTGTCTCCAG 4507
Qy 5065 gectacccttgctcaggttcaagaagacagtgaaagagctcatccgcactctctttct 5124
Db 4508 GCGCGCGCTTGCTGAGGTGAAGAGATAGTGAGAGGCTCGCTGTCATCTGCTTCT 4567
Qy 5125 ttgcagaaattctcaatacaaatgagtgctccattgggtgtgatattcctgctggtatt 5184
Db 4568 TTGTGGAAGTGTGCGAGCACACAGACGGCTCAGTGGGTGCGGTGCTCCCGGCTGGTACC 4627
Qy 5185 ttgggtctcactgaagaatggcatcaccccttactctgctgtgctgcatgtttttcc 5244
Db 4628 TGTGCTCTCCCTGAGGAACGGCATCACCCCTTACTCCGCTGTGCTGCACTGCTTTTCC 4687
Qy 5245 actattacttgggttaactccgctgaggaactgcataccaattctgcagaagagagt 5304
Db 4688 ACTATTTACTTGGAGTAGCTCGGCTGAAGAACTGTTGCCAATCTGCTGAAGAGAAAT 4747
Qy 5305 acagtgcactctgtagctatctatttacctacaaaattgttctcctgctctccaggaat 5364
Db 4748 TCAGTGCACCTCTGTAGCTATCTATTACCCACAAAATTTGTTCTGCTTTTCCAGGAAT 4807
Qy 5365 attgggatactgaagcccttgctccagagtggtgtgcagatacctgcttactaaact 5424
Db 4808 ATTGGGATACCAATAGGCCCTTACTACAGAGGTGTGTGGAGATCCTGCTTACTCAAGT 4867
Qy 5425 gttgaagcaaaaaaacccgtggtcaggtaccctlagaaaaaagaatagtgtgatagac 5484
Db 4868 CTTTGAAGCAGAAAAGTCTGTGTGTCAGTACCCCTAGAAAAAGAAATAGTTTGTATAGAC 4927
Qy 5485 ttctgtgactgactagctcctctgaatcaagcttctcattctcaggtgcccacggtctg 5544
Db 4928 TTCTCTGAGCACTACAGCTGTCTCTTAATCAGGCTTCTCACTTAGTGTCTCCAGGCTG 4987
Qy 5545 cagatgtagcgaagaagcatcgtccctcctcctctctctctctctctctctctctctctc 5604
Db 4988 CAGATGATGAGCAAAAGCATCTGTCCTCTGCTTTTCTGTGGGCCATCTGCTGTCTC 5047
Qy 5605 aqaacattgctgccagaaaattggaacgggaagaggttgagagcttcacattttcacg 5664
Db 5048 AGAATATCTGTTGCCAGAAATAGTGAATGGGGAAGAGTTTGAGCGTCCGTTTTTCATG 5107
Qy 5665 cacttcactgtgagccggagctgctcattttctctaaaaatcagaaatccgagtggtcc 5724
Db 5108 CGCTTCATTTGGTCTGGAGTCTCAATTTCTTAAAAATCCGAGAAATCCAGGTTGTTCC 5167
Qy 5725 tgggtgaaggttaaacgagagctgtgctctccagctcctcacttggtgtaataatggag 5784
Db 5168 TGGTGGAAAGAAAACCCAGAGCTGTGCTACCCGCTTACTTGGATGAATATGGAG 5227
Qy 5785 aaacagacccctgctgaagaggggcaacccctcatttactctcgtgagcgtatcgga 5844
Db 5228 AAACAGACCCAGGGCTAAAGAGAGAAACCCACTTCAATTTACTCGGAGCGGTATCGGA 5287
Qy 5845 agctccatttggctggaacaaacactgcattatagaagagattgctagagccagagaga 5904
Db 5288 AGCTCATTTGGTCTGGCAACAGCACTGCATTTATAGAAGAGATTGCTCGAGCCAGGAGA 5347
Qy 5905 ctaatcagatgttatgttgatccaactggcaggttactgtgagctccaactcgtcctcaag 5964
Db 5348 CTAATCAGATGCTATTGGAATTAACCTGGCAGTTACTTCTGAGCTTCAGTTCTGCCTCAAG 5407
Qy 5965 acaatcacaaatgacagacagtagtaagcgtgattcaaaaattatggaaaaacittctgagg 6024
Db 5408 ACAATCATGAGTGACATCAATAATAAAGACTGATCTAAATTTCTAGAGAACTTTCTGAGG 5467
Qy 6025 gctgggaagattgagggtcttttgcctcctgctccatgctccagaggttcacttacatcaataa 6084

Db 5468 ACGGGGAAAGTATTGGAGGTCTTTTGTATCCATGTCAGATTCACACACATTAATAAAAT 5527
Qy 6085 attcttaaggagattgctttcaatagcaaacatagcttcacagaaaaa-aggac 6143
Db 5528 ATTCTTAATGGAATATTGCTTCAATATCAACATTAAGCTTCAAGGAAAAACAAGAC 5587
Qy 6144 atagatcaatctgt-----ttt 6160
Db 5588 ATAGATTAAATGTTTATGTTCTAGAACACTAAAGAAATGCTTGTTCATCCAAGTGTCTAT 5647
Qy 6161 atgctgtagtattccaggaaatttattcccttccataatttgctcatttctattttatt 6220
Db 5648 TTCTGCTAATATTTCAGAAAACTCTTCCCTTTCATAAAGTCTCTAGTTTCAATTTCAAT 5707
Qy 6221 catccactgtagatgaatgcacgtcaaacagctgttagacattttatgtgttgtaac 6280
Db 5708 CACCCACTGGTAAATGAGGTCACTAAGCAATTTGGACATTTCTCCATCTGGCTAAC 5767
Qy 6281 tctctgcaattttgtatttggtgttt 6307
Db 5768 ATCTCTGCACCTTTGTATTGGTGT 5794

RESULT 3
US-08-982-956-2
; Sequence 2, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..999
US-08-982-956-2

Query Match 15.8%; Score 996.2; DB 2; Length 1001;
Best Local Similarity 99.7%; Pred. No. 7.6e-279;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2799 atggatccccataaagtctgttactggtacttcagaggtatgaacttgccagagctttt 2858

Db 421 CAGCATACGAGGCTGAACATATGACAGAGAAAGAGAGAAACAAAGATGAA 480
Qy 3279 gcattgcgcacacaccacctcctgaattctgcctcgtttcagcaaaagtattacactt 3338
Db 481 GCATTGCCGCCACACACCCTCTGAAATTCGCCCTGCTTCAGCAAGATGATTAACCTT 540
Qy 3339 ctcaactgtatcatgatgatcatctcaggaccgtatttgagcggcaatagacaca 3398
Db 541 CTCAACTGTGATATCATGATGATGATCTCAGGACCGTATTGAGCGGCAATAACACA 600
Qy 3399 gatttcaacttggaccgaaggatgctcaaatggcttttcatcttcttgcattgggt 3458
Db 601 GATTCTAACTTGTGACCGCAAGGATGCTCCAATGGCTTTTCATATTTCTGCGCATGGGT 660
Qy 3459 ttactagaagaagcaacagcttcaaaaagctcctgaagaagaagtaacatttgacttt 3518
Db 661 TTACTAGAGAGAGACACAGCTTCAAAAAGCTCTGAAGAGAAAGTAACTTTGACTTT 720
Qy 3519 tatcataaggcttcaagattgggaagttcagccatgaataatacaaatgcttttggaaaaa 3578
Db 721 TATCATAAGGCTTCAAGATTGGGAAGTTTCAGCCATGATATACAAATGCTTTTGGAAAA 780
Qy 3579 ctcaagaagaattccccagttgaaggccagagagacatgataaocgtgatacttcagatg 3638
Db 781 CTCGAAGAAATTCGCCAGTTAGAGGCCAGAGGACATGATAACGTGGATACTTTCAGATG 840
Qy 3639 tttaacacagtgagcagattgaagaaaaatctgtttaattgtagcaaccacatcagga 3698
Db 841 TTTGACACAGTGAAGCGATTAGAGAAAAATCTTGTAAATGTAGCAACACATCAGGA 900
Qy 3699 tcggaatctattagaatgatgagattactcatgatataaagaaagcagaaagaaaga 3758
Db 901 TCGGAATCTATTAGAATGATGAGATTACTCATGATAAGAAAAAGCAGAGAAAAAGA 960
Qy 3759 aaagctgaagctgtaggctacatgcgcagaaagatcggc 3799
Db 961 AAAGCTGAAGCTGTAGGCTTCATCGCAGAGATCATGGC 1001

RESULT 5

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 0.9%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 9.4%; Pred. NO. 6e-06;
Matches 43; Conservative 219; Mismatches 195; Indels 0; Gaps 0;

Qy 2862 aagaccatatctcaaaaagaccaggatttgattaaacaataataacataaagaaga 2921
Db 1481 AATTACTTATCTATGCAAGTAGTTAAAGAGATAGAAATTTGGTACRRRRRRRRRR 1422
Qy 2922 atgtctcaggtcctcatctattgtgggtgagcgttatgtacctgagtgaggaaatgtg 2981
Db 1421 RRR 1362
Qy 2982 accaaagaagagtgacacatgagagaatcattcattgcttgcattgaacccatgcca 3041
Db 1361 RRR 1302
Qy 3042 cacagtgccattgcccaaaattacctgagaatgaaataatgaaactggcttagagaat 3101
Db 1301 RRR 1242
Qy 3102 gtcataaacaagtgccacatttaagaacacagtgatcatcagccatggagttatgaa 3161
Db 1241 RRR 1182
Qy 3162 ctaaaagatgaatcactgaagacattcaatatgtacttttattctatctcaaaacccag 3221
Db 1181 RRR 1122
Qy 3222 catagcaagctgaacatgacagagaagaagagagaaacaaagaagaagatgaaagca 3281
Db 1121 RRR 1062
Qy 3282 ttgcccaccaccacctctcgtgaattctgcctgctt 3318
Db 1061 AGCTCCCTCGACCTGCAGCCAAAGCTCGGAATTAATT 1025

RESULT 6

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk


```

; TELEFAX: 650-327-3231
;
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-009-913-1
;
; Query Match 0.7%; Score 44.2; DB 3; Length 72928;
; Best Local Similarity 86.0%; Pred.No.0.18;
; Matches 49; Conservative 0; Mismatches 8; Indels 0; Gaps
;
QY 303 cccctgaaacacacattctcttctgtctatgaatttgactactctagctggat 359
Db 19091 CCCCGGCAACCACCATTCTACTTCTGTCTCTATGAATTTGACCACCTCTAAGTAGCT 19035
;
RESULT 9
US-08-991-789A-236/C
; Sequence 236, Application US/08991789A
; Patent No. 6225034
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; Smith, John M.
; Reed, Steven G.
;
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
;
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Ip Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:
;
US-08-991-789A-236
;
; Query Match 0.7%; Score 42.2; DB 4; Length 325;
; Best Local Similarity 79.4%; Pred.No.0.021;
; Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps
;
QY 298 acacacccctggaaacacacattctacttctgtgtctatgaatttgactactctagctgg 351

```

```

Db 283 AAAAGCCTGGCAACACCACTTATTGTGTCTATGAAATTCACCTACTAGGTAC 224
Qy 358 atc 360
Db 223 CTC 221

RESULT 10
US-08-963-602-6/c
: Sequence 6, Application US/08963602
: Patent No. 6090554
: GENERAL INFORMATION:
: APPLICANT: Woychik, Richard
: APPLICANT: Garfinkel, David
: TITLE OF INVENTION: EFFICIENT CONSTRUCTION OF GENE
: TITLE OF INVENTION: TARGETING VECTORS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
: STREET: 233 South Wacker Drive/6300 Sears Tower
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/963,602
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: REFERENCE/DOCKET NUMBER: 01017/33985
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 474-6300
: TELEFAX: (312) 474-0448
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1517 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "primer"
US-08-963-602-6

Query Match 0.6%; Score 41; DB 3; Length 1517;
Best Local Similarity 56.2%; Pred. No. 0.13;
Matches 77; Conservative 0; Mismatches 60; Indels 0; Gaps

Qy 109 gaaacattacctcaaggcttatcatctgttttccacttatccacaagctgctatgac 168
Db 1036 GTGTGCTTGGCTTCTGAGAGCCGCTGTCTTCTTGGGGCGCTGGTACCGCGCTGCTGCTGCTG 977

Qy 169 cactgctgtgcccgaacctggaaccacccgagcccccactactgctccactaccactg 228
Db 976 AGTCCCGCTGCCGCGCTGCTGCCACCACCAATGTACTGTACTGCGCGCTCCACCACTGC 917

Qy 229 ttctccaccctgatca 245
Db 916 TGCCTCTCTCTCCACCA 900

RESULT 11
US-09-128-155-16
: Sequence 16, Application US/09128155
: Patent No. 6117654

```



```

; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE OF INVENTION: AND USES THEREOF
; TITLE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

	Query Match	0.6%	Score 40.4;	DB 3;	Length 152331;
	Best Local Similarity	81.0%;	Pred. No. 3.6;		
	Matches 47; Conservative	0;	Mismatches 11;	Indels 0;	Gaps 0;
Oy	299	cacacctggaaaccaccattcttctgtgctcgaatttgactactctagctg	356		
Dd	85997	caggcccgagcaatcacattctagtttctgcctcatgatgttcgacaacttaggtq	86054		

```

RESULT 12
US-08-299-953-1/c
; Sequence 1, Application US/08299953
; Patent No. 5646333
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,953
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

```

US-08-299-953-1

Query Match          0.6%; Score 39.8; DB 1; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 3607 agaaggacatgataaacgtggatacttcagatgtttgacacagtgaaagcattaaagaaa 3666
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 AAAAGAAGACGGAGGAGTAAATATTTTAGATTTTTCACGTATTTAAAAAATTATATCAA 1231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3667 aatctgtttaattgtagcaaccacatcagatcggaatctattaaagaatgatgagatta 3726
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 AAAAATATCTTTATATATTAATAAATGGAGAAAGATAATTTATATTAATTAANAANAAG 1171
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3727 ctcatgataaagaaaaagcagaaacgaaagaaagctggaagctgctaggctacatcgcc 3786
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1170 ATAATAATAAAATCTAGAGTTATAATAAANAACATAATATTAATTCITTTTACATTTGA 1111
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3787 agaagataatggtccagatgtctgccttacagaaaacttcattgaacctcaataactca 3846
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1110 AAATGATTTTATTTATGATATAATTTTTTTTCAAAACCAACCAATAATAAANAATGATAGG 1051
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3847 tgtatgacaatacatacagaatgcctgggaaagaa 3881
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 AGTATTATCATATGCTCAGAAATTTATTAAAGAAGAA 1016
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-08-459-415-1/c
; Sequence 1, Application US/08459415
; Patent No. 5744334
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,415
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-459-415-1

```

Query Match	0.6%	Score 39.8	DB 1	Length 2861
Best Local Similarity	46.5%	Pred. No. 0.42		
Matches 128	Conservative	0	Mismatches 147	Indels 0
QY 3607	agaaggacatgataa	cgtggataacttcagat	gttttgacacagtcgaagcgattaaagaaa	3666
Db 1290	AAAAGAACGCGGAG	GTTAATATTTTAGAT	TTTTTACACGTATTAAAAAAATTTATATCAA	1231
QY 3667	aatctgttcaattgt	tagcaaccacatcagat	cggaaatctattaaagaatgatgattca	3726
Db 1230	AAAAATATCTTTAT	ATATTAATGGAGAGATAA	TTTATATTAATTAATAAAAAAAG	1171
QY 3727	ctcatgataaagaaa	agcgagaaagaaagct	gaagctgctaggtcacatcgcc	3786
Db 1170	ATAATATAAATCT	AGAGTTATAATTAANAAC	TAAATTAATTCCTTTTAACATTGTA	1111
QY 3787	agaagatcatggtcg	atgctgtccttacagaaaacttc	atttgaaactcataaactca	3846
Db 1110	AAATGATTTTATAT	GCATATATAATTTTTTTC	AAAACAACCATATAAAAAATGATAGG	1051
QY 3847	tgtatgacatacatc	agaaaatgcctgggaaaga		3881
Db 1050	AGTATTATCATATG	CTCAGAAATTTATTATAA	AAAGAA	1016

```

RESULT 14
PCT-US95-11231-1/c
; Sequence 1, Application PC/TUS9511231
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnrur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
; OF FOREIGN PROTEINS TO THE PLANT EPIDERMIS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11231

```

Query Match 0.6%; Score 39.8; DB 5; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;

```

Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
Qy 3607 aagaggacatgataacgctggatacttcagatgtttgacacagtgaaagcgattaaagaaa 3666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 AAAAAAGACGAGGAGGAGTAAATATTTTAGATTTTTCACAGCTATTAAAAAAATATATCAA 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3667 aatctgtttaattgttagcaaccacatcagagatcggaatctctattaagaagtatgagatta 3726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 AAAAATATCTTTATATATTAATAAATGGAGAGATAAATTTATATTAATTAATAAAAAAAG 1171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3727 ctcatgataaagaaaaagcagaacgaaaaaagaaagctgaagctgctgagctacatcgcc 3786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1170 ATAATAATAAAATCTAGAGTTAATAAATAAAACATAATTAATCTTTTAACATTTGTA 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3787 aagaatcatgctcgcatgctgctctacagaaaaacttcattgaaactcataaactca 3846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1110 AAATGATTTATATATGATATAATTTTTTTCAAAAACAACATAAATAAAAAATCATAGG 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3847 tgatgacaatcacatcagaatgcctgggaaagaa 3881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 AGTATTATCATCTCAGAAATTTATTAAGAA 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-08-299-953-2/c
; Sequence 2, Application US/08299953
; Patent No. 5646333
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur.
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/299,953
; APPLICATION NUMBER: US/08/299,953
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-299-953-2

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:32:35 ; Search time 2649.68 Seconds
(without alignments)
25582.097 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattgcgcagagg.....aatttattgttggtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	674.6	10.7	682	10	AI929033	AI929033 au64c10.y
2	656.6	10.4	756	11	BG334574	BG334574 602533425
3	637.4	10.1	641	10	AI361043	AI361043 qy03f11.x
4	627.2	9.9	797	11	BG862813	BG862813 602799074
5	573.2	9.1	782	11	BI086469	BI086469 602849734
6	557	8.8	565	11	BF063405	BF063405 7b89d06.x
7	512.2	8.1	606	10	BE589438	BE589438 195602 BA
8	479.8	7.6	624	11	BG219270	BG219270 RST39023
9	477.8	7.6	481	10	AI192195	AI192195 qc92608.x
10	473.4	7.5	583	10	AW971391	AW971391 EST383480
11	471.6	7.5	478	10	AA401319	AA401319 zu63d04.r
12	454	7.2	522	10	AW291190	AW291190 UI-H-BI2-

13	449	7.1	522	11	BG382648	BG382648 298222 MA
14	448	7.1	520	11	BG382624	BG382624 298590 MA
15	410.8	6.5	486	10	BE650873	BE650873 UI-M-BH3-
16	354.8	5.6	925	10	BE573578	BE573578 601333605
17	352	5.6	374	10	AA400279	AA400279 zu63d04.s
18	351.4	5.6	435	10	AW489271	AW489271 UI-M-BH3-
19	330	5.2	394	11	BF774117	BF774117 283680 MA
20	318	5.0	481	10	AI187306	AI187306 qf28h08.x
21	307.2	4.9	397	10	AW311960	AW311960 6005 MARC
22	301.8	4.8	504	10	AI980640	AI980640 pat_pk002
23	266.2	4.2	327	10	BE654236	BE654236 UI-M-AKI-
24	265.2	4.2	523	11	BG797647	BG797647 lcl16b02.x
25	263.4	4.2	300	11	Z17892	Z17892 HSDHE1041 S
26	252.6	4.0	1079	11	BG292980	BG292980 602389655
27	249	3.9	562	10	AI504731	AI504731 v113c11.x
28	243.6	3.9	372	11	BF542537	BF542537 UI-R-C3-s
29	242.8	3.8	505	11	BG364916	BG364916 dc93c08.y
30	241	3.8	325	10	AI615529	AI615529 v113c11.y
31	237	3.8	291	10	AV225341	AV225341 AV225341
32	235.8	3.7	396	11	BE930958	BE930958 RC3-GN007
33	232.8	3.7	256	10	AI921294	AI921294 w023a03.x
34	231.4	3.7	278	10	BE077143	BE077143 RC5-BT060
35	230.6	3.7	323	11	H33916	H33916 EST110358 R
36	223.6	3.5	745	10	AW976158	AW976158 EST388267
37	222	3.5	555	11	BE873236	BE873236 601451771
38	217	3.4	685	10	AI693180	AI693180 w068e01.x
39	209	3.3	313	10	AV168252	AV168252 AV168252
40	208.6	3.3	632	11	BG625558	BG625558 pgn1c.pk0
41	206.8	3.3	221	10	AA507138	AA507138 nh42d07.s
42	204	3.2	1148	11	BF164318	BF164318 601773090
43	201	3.2	620	10	AW702134	AW702134 uq98d05.x
44	200.4	3.2	435	11	BE930879	BE930879 RC3-GN007
45	200	3.2	711	10	AI646734	AI646734 ub65b06.x

ALIGNMENTS

RESULT 1
AI929033 682 bp mRNA EST 23-AUG-1999
LOCUS au64c10.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION IMAGE:2519538 5' similar to TR:O70481 O70481 UBIQUITIN-PROTEIN
LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.
ACCESSION AI929033 GI:5664997
VERSION AI929033.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 682)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubouque, T., Geisels, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Other_ESTS: au64c10.x1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 450.
Location/Qualifiers
1..682
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="IMAGE:2519538"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site.1: SstI; Site.2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGATCTTAAATTAATCCCTCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGCTGAGATCTTAAATTAATCCCTCCCCCCCC-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
BASE COUNT 178 a 155 c 166 g 182 t 1 others
ORIGIN

Query Match 10.7%; Score 674.6; DB 10; Length 682;
Best Local Similarity 99.3%; Pred. No. 3.9e-147;
Matches 677; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 5261 aactcgcctgaggaactcatcaccatctgcagagagagatcagtcactctgtag 5320
DB 1 AACTCGCCTGAGGAAGTGCATACCAATCTGCAGAGAGAGAGTACAGTGCACCTGTAG 60
QY 5321 ctatctatcttacctacaaaattgtctcgtctctccagggaattgggatactgaag 5380
DB 61 CTATCTATCTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATGGATACTGTAG 120
QY 5381 gccctgtctccagaggggtgctgagatcctccttactactaaactgtttgaagcaaaaaa 5440
DB 121 GCCCTTGCTCCAGAGAGTGGTGTGCAGATCTGCTTACTAACTGTTTGAAGCAAAAAA 180
QY 5441 caccctgtcaggtacccttagaaaaaataagtttgatagagcttccgtgactatag 5500
DB 181 CACCGTGGTCCAGTACCCCTAGAAAAAGAAATAGTTTGATAGAGCTTCCGTGATGACTATAG 240
QY 5501 ctgctctcctgaatcaagctctcatttcagggtgcccacggtctgcagatgagagcaaa 5560
DB 241 CTGCTCTCTGATCAGCTTCTCAATTCAGGTGCCACCGTCTGCAGATGATGAGCGCAA 300
QY 5561 gcaatcgtctcctgtcttctgtggtggtatatactatgttctcagaacatttgcgtgcca 5620
DB 301 GCATCCTGCTCTGCTCTTCTGTGGGGCTATACTATGTTCTCAGAACATTTGCTGCCA 360
QY 5621 ggaattgtgaacggggaagaggttgagcttgcatttttcacgcaacttcaactgtgagc 5680
DB 361 GGAATTTGTGAACGGGGAAGAGTGGAGCTTGCATTTTTCACGCACTTCACTGTGGAGC 420
QY 5681 cggagctctgctatttctctaaaaatcagagaatgcccagtggtccttgggtgaaggtaaagc 5740
DB 421 CGAGTCTGCAATTTTCTTAAAAATCAGAAATGCCGAGTGGTCTCGTGTGAAGGTAAAGC 480
QY 5741 cagaggtctgtgctatccagctccttacttggatgaataatggagaaacagaccctggcct 5800
DB 481 CAGAGGCTGTGCTATCCAGCTCCTTACTTGGATGAATATGAGAAACAGACCCCTGGCCT 540
QY 5801 gaagggggcaaccccccttcattatctcgtgagcgggtatcggagagctccatttggctctg 5860
DB 541 GAAGAGGGGCAACCCCTTCAATTTATCTCGTGAGCGGTATCGGAAGCTCCATTTGGTCTG 600
QY 5861 gcaacaacactgcattatagaagagatgctaggagccaaagagactaatcagatgttatt 5920
DB 601 GCAACAACACTGCATATAGAAGAGATTGCTAGGAGCCAAAGAGACTAATCAGATGTATT 660

QY 5921 tggattcaactggcaggttactg 5942
||||| | |||| ||||
DB 661 GGGATTCCCTGCAGNACTG 682
RESULT 2
BG534574 756 bp mRNA 03-APR-2001
LOCUS 602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
DEFINITION mRNA sequence.
ACCESSION BG534574
VERSION BG534574.1 GI:13526116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 756)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM1465 row: a column: 07
High quality sequence stop: 751.
Location/Qualifiers
1. 756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4663182"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site.1:
SfiI (ggcgctcgcc); Site.2: SfiI (ggcattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATTTATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGGCAGATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 274 a 138 c 152 g 192 t
ORIGIN

Query Match 10.4%; Score 656.6; DB 11; Length 756;
Best Local Similarity 95.9%; Pred. No. 6.5e-143;
Matches 730; Conservative 0; Mismatches 19; Indels 12; Gaps 5;
QY 2836 ggtatgaactgcagagcttttaacaagaccatctacaaagaccaggtattgatta 2895
DB 1 GGTATGAACCTGGCGAGGCTTTTAAACAGACCATCTACAAAAGACCAGGATTTGATTA 60
QY 2896 acaataataactaactaagaagaatgcttcagggtcctctatctatattgtgggtgagc 2955
DB 61 AACAAATATACTACTAATAGAAATGCTTCAGGTCTCTATATTGTTGGGTGAGC 120
QY 2956 gttatgtacctgagtgagggaatgtgaccaagaagaggtgcacatgagagaaatcattc 3015
DB 121 GTTATGTACCTGGAGTGGGAAATGTCACCAAGAGAGGTGCAATGAGAGAAATCATTC 180
QY 3016 acttgtttgcatggaaccatgcacacagtgccattgcccataaatttacctgagagtg 3075
DB 181 ACTTGCTTTTCATTGAACCCATGCACACAGTGCATTGCCAAAAATTTTACCTGAGAATG 240

Qy	3076	aaaaataatgaactgcttagaagaatgtcatataaacaagaatggtccacattttaagaacacag	3135
Db	241	AAAAATAATGAACATGGCTTAGACAATGTCTATAAACAAAGTGGCCACATTTAAGAAACACG	300
Qy	3136	gtgtacagccatgaggttatgaactaaaaatgaataactgaagacttcaatatgt	3195
Db	301	GTGTATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACAATGAAAGACTTCAATATGT	360
Qy	3196	actttatcattactccaaacccagcagcagaggtgcaacatatagcagaagaaaagga	3255
Db	361	ACTTTTATCATTAACCTCAAACCCAGCATAGCAAGGCTGAACATATGCCAGAAAGAAAGGA	420
Qy	3256	gaaacaagaataaagaatgaag-----catggccgcacacccacctcctgaattct	3309
Db	421	GAATAACAAGAAAAAACAAGATGAAGGTAAAAACATATGCCGACCAACACCTCTCTGAATTCT	480
Qy	3310	gcctgtttcagcaaaagtattaaaccttctcaactgtgatat-caatgatgcattcttc	3368
Db	481	GCCCTGCTTTTCAGCAAAAGTATTAACTTCTCAACTGTGATATCCATGATGTACATTCTC	540
Qy	3369	aggaccgtatttgagcgggcaatagacacagattctaaacttggaccgaagggatgctc	3428
Db	541	AGGACCGTATTTCAGCGGGCAATAGACACAGATTTCTAACTTGTGGACCAAGGATGCTC	600
Qy	3429	caaatggcttttcatactctggcattgggtttactagaagagagcaacagcttcaaaaa	3488
Db	601	CAAAATGGTTTTCATATTCTTCGCAATTTGGTTTTACTAGAGAGAAGCAACAG-TTCAAAAA	659
Qy	3489	gctcctgaagaagaagtaacattgacttttcatataagagcttcaagattgggaagtcca	3548
Db	660	GCTCCTCAGAGAAGAAGTAACATTTGACTTTTATCATTAAGGCTTTCAAGATTGGGAAGTTCA	719
Qy	3549	gccatgaatacaaatgctcttttggaaaaactcaaaaggaat	3589
Db	720	G-CATGATATACAA---TGCTTTGGAAAAAUAAGAGATT	756
RESULT	3		
AI361043/C			
LOCUS	AI361043	641 bp	EST
DEFINITION	qy03f11.x1 NCI-CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2010957 3'		15-FEB-1999
	Similar to TR:P91133 P91133 SIMILAR TOS. CEREBISIAE		
	UBIQUITIN-PROTEIN LIGASE E3 COMPONENT SP:P19812. ; mRNA sequence.		
ACCESSION	AI361043		
VERSION	AI361043.1	GI:4112664	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 641)		
AUTHORS	NCI/NIH/NC-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index		
JOURNAL	Unpublished (1998)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 1083 Seq primer: -400P from gibco High quality sequence stop: 453. Location/Qualifiers 1..641 /organism="Homo sapiens"		
FEATURES	source		

Query Match	10.1%	Score 637.4	DB 10	Length 641
Best Local Similarity	99.5%	Pred. No. 2e-138		
Matches 638	Conservative 0	Mismatches 3	Indels 0	Gaps 0
Qy 5457	cctgagaaaaaataagtttgaatagagcttccctgaactatactgacctgaatcaa	5516		
Db 641	CCTGAGAAAAAATAGTTTGTATAGAGTTCCTGATGACTATAGCTGCCTNTGAATCCA	582		
Qy 5517	gcttctcaattcaggtgccacgctctgcaatgatagcgaagaactcctgctcctgc	5576		
Db 581	GCTTCTCATNTACAGTGGCCACGGTCTGCAGATGATGAGCAAGACATCTGCTCCTCTGC	522		
Qy 5577	cttttctgtggggtatactatgttctcagaacatttgcgtccaggaaattgtgaacggg	5636		
Db 521	CTTTCTGTGGGCTATACTATGTCTCAGAAACATTTGTCGCCAGGAATTTGTGAACGGG	462		
Qy 5637	gaagaggttgagcttgcaatttccaagcacttcactgtggagccgagctgcaatttc	5696		
Db 461	GAAGAGGTTGAGGCTTGCAATTTTTCACGCACCTTCTGAGGCGGAGTCTGCATTTTC	402		
Qy 5697	ctaaaaatcagaaatgccgagtgctgttgaagtaaaagccagagctgtgcctat	5756		
Db 401	CTAAAAATCAGAAATGCCGAGTGGTCTGTTGAAGTAAAGCCAGAGGCTGTGCCTAT	342		
Qy 5757	ccagctcctacttgatgaatatggagaaacagacctgacctgaagagggggcaacccc	5816		
Db 341	CCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCTGSCCTGAAGAGGGGCAACCCC	282		
Qy 5817	ctcatttatctgtgagcgttatcggaagctccatttgcctggcaacaacactgatt	5876		
Db 281	CTTCATTTATCTCGTGAGCGGTATCGGAAGTCCCACTTTGGTCTGGCAACACACTGCATT	222		
Qy 5877	atagaagagattgtaggagccaagagaactaatcagatgttatttgattcaactggcag	5936		
Db 221	ATAGAAGAGATTGCTAGGAGCCAGAGACTAATCAGATGTTATTTGGATTCACTGGCAG	162		
Qy 5937	ttactgtgagctccaaactctgcctcaagacaatcacaaatgacacagtagtaaaagctg	5996		
Db 161	TTACTGTGAGCTCCAACCTCGCTCAAGACAAATCACAAATCAGCAGTAGTAAGAGCTG	102		
Qy 5997	attcaaaaattatggaaaaacttctgagggctgggaagtatgtgagggcttttgcctcca	6056		
Db 101	ATTCAAAATTTATGGAAAACTTTCTGAGGCTGGGAAAGTATTGGAGGGCTCTTTTGCTCCA	42		
Qy 6057	tgtccaggttcaactacatcaataaaatattttcctaattgga	6097		
Db 41	TGTCCAGGTTTCACTTACATCAATAAAATATTTCCTTAATGGA	1		

BASE COUNT

174 a 155 c 135 g 175 t 2 others

ORIGIN

db_xref="taxon:9606"

/clone="IMAGE:2010957"

/clone_lib="NCI_CGAP_Brn23"

/tissue_type="glioblastoma (pooled)"

/lab_host="DH10B"

/note="organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTCAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

2010957

NCI_CGAP_Brn23

glioblastoma (pooled)

DH10B

organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTCAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

RESULT 4

LOCUS BG862813

DEFINITION 602799074P1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4934370 5', mRNA sequence.

ACCESSION BG862813

94 MAY 2001

EST

29-MAY-2001

IMAGE:4934370 5',

Qy	5591	tatactatgtttctcagaacatttctgtccagaaattgtgaacggggaagaggtlggagc	5655
Db	422	CATCCTGTGTTCTCAGAACATCTGTGTGCCAAGAAATAGTGAATGGGGAAGAGGTTGGAGC	481
Qy	5651	ttgcatttttcacgcacctctcactgtggagccgagtcgcattttctctaaataacagaga	5710
Db	482	GTGCGGTTTTTCATGCGCTTCATTTGGTGCTGGAGTCTGCATTTTCTCTAAAAATCCGAGA	541
Qy	5711	atgccgagtggtccttggttgaaggtaaaagccagagcgtgtgcctatccagctccttactt	5770
Db	542	ATGCAGGGTGGTCTCGTGGTGAAGAAAGCCAGAGGCTCTGCCTACCCAGCCCCCTTACTT	601
Qy	5771	ggatgaatatggagaaacagacccctggcctgaaggggcaacccctcttattatctcg	5830
Db	602	GGATGAATATGGAGAAACAGACCCAGGGCTAAAGAGAGAGAAACCCACTTCATTATCTCG	661
Qy	5831	tgagcggatcggaagctccatttgctc-tggcaacacacitgcattatagaagaattg	5889
Db	662	GGAGCGGTATCGGAAGCTGCATTTGGTCTTGGCAAAAGACTTGGATTTATAGAAGAATG	721
Qy	5890	ctaggagccaagagactaatcagatgttatttggattcaactggcagtgactgtgagctc	5949
Db	722	GTCGAGCGCGGAGAGACTAATCAGATGCTATTTGGACTTAACCTGGGAGTTACTCTGAGATC	781
Qy	5950	caactctgcctcaaga	5965
Db	782	CAGTTGTCTCAAGAGA	797
RESULT 5			
BI086469			
LOCUS	602849734P1	782 bp	mRNA
DEFINITION	Homo sapiens cDNA clone IMAGE:4991193 5'		
ACCESSION	BI086469		
VERSION	BI086469.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 782)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Incyte Genomics, Inc. cDNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11008 row: d column: 10 High quality sequence stop: 715. Location/Qualifiers 1. 782 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4991193" /clone_lib="NIH_MGC_10" /cell_line="MGC36" /lab_host="DH10B" /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: Not Site 2: SalI; Cloned unidirectionally. Primer: Oligo Average insert size 1.5 kb. Library prepared by Life Technologies."		
BASE COUNT	184 a	194 c	163 g
ORIGIN	241 t		

```
Query Match          9.1%; Score 573.2; DB 11; Length 782;
Best Local Similarity 92.5%; Pred. No. 2.2e-123;
Matches 712; Conservative 0; Mismatches 48; Indels 10; Gaps 10;

QY 4863 aaatcagaagatacacaccgctctgtctctatagatctgtttcatgttttgggtgct 4922
|||||
Db 4 AAATCAGAGATACACCATG-CTTCTGCTATAGATCTGTTACATG-TTTGGTGGATGCT 61
|||||

QY 4923 gtgttagcatcccatctctgttattggatgacccctgttgaactcgcagcctcttcagtt 4982
|||||
Db 62 GTGTTAGCA-TCCCATCTCTGTA-TGGGATGACCCCTGTTGATCTGCAGCCTCTTCAGTT 119
|||||

QY 4983 agttctctcataaacacacttattctctccattgtatccatggcacacatgcttcag 5042
|||||
Db 120 AGTTCTCTCTATAACCACTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCAG 179
|||||

QY 5043 atactactacagttagacacagcctacccctgtctcaggttccaagaagacagtgaag 5102
|||||
Db 180 ATACTACTTACAGTAGACACAGCGCTACCCCTTGCTCAGGTTCAAGAGACACAGTGAAGAG 239
|||||

QY 5103 gctcattccgcatctcttcttgcagaaattctcctaatacagaagtggtcccatggg 5162
|||||
Db 240 GCTCATTCGCACTCTCTCTTCGCAAGAAATTTCTCAATATACAAAGTGGCTCCATTTGG 299
|||||

QY 5163 tgtgatattcctggctggtgatttgggtctcactgaagaatgg-catacacccttatct 5221
|||||
Db 300 TGTGATATTCTGGCTGCTGATTTGTTGGTCTCTCACTGAAGATGGCCATCACCCCTTATCT 359
|||||

QY 5222 tgcgtgctgcattggttttccacatttacttgggttaactcgcctgcgtaagaactgca 5281
|||||
Db 360 TCCTGTCTGCTGATTTGTTTTCCTACTATTTACTTGGGGTAACTCCGCGCTGAGGAACCTGA 419
|||||

QY 5282 taccattctcagaagagagatgaca-gtgcactctgtagctatctatttacctacaa 5340
|||||
Db 420 TACCAATCTGCAGAAGAGAGATGACACGTGCACCTCTGTAGCTATCTATCTTTTACCTACAA 479
|||||

QY 5341 attgttctctctctccaggaaattgggatactgtgaagccctgtctccagagtggt 5400
|||||
Db 480 ATTTGTTCTCTCTCTCCAGGAATATTGGGATACCTGTAAGGCCCTTTTTCAGAGGTGCT 539
|||||

QY 5401 gtgca-gatcctgcttactaa-acgtgttgaagcaaaaaaacacccgtggta-ggtacc 5457
|||||
Db 540 GTGAGGATCTCGCTTACTAACACTGTTTGAACGCAAAAACACGCTGGTCAGGCTACC 599
|||||

QY 5458 ctgaaaaaagaatggtttagagcttctgtagctatagctgctcctcctgaatca-a 5516
|||||
Db 600 CTAGACAGACATAGTTTGTATAGAGCTTCCCTGATGATATAGCTGCTCTGCTGAAATACA 659
|||||

QY 5517 gcttctcattcagtgccacggtctgcagatgatgagcaaaagcatacctgtcctctgc 5576
|||||
Db 660 GCTTCTCATCTCAGTGCCCGCTGCTGCAGATGATTGACCCAAAGCATCTCTGCTCCGC 719
|||||

QY 5577 cttttctgtgggctatactattgttctcagaacattgtgcacaggaat 5626
|||||
Db 720 TCTTCTGTGGGCTATAACTATGCTCCGAACCACTTTGCTGCAGAAAT 769
|||||

RESULT 6
BF063405/c
LOCUS
DEFINITION
N-RECQGNIN ; mRNA sequence.
16-OCT-2000
7h89d06.x1 NCI-CCAP.Col16 Homo sapiens cDNA clone IMAGE:3323147 3'
similar to TR:070481 070481 UBIQUITIN-PROTEIN LIGASE E3 COMPONEN
ACCESSION BF063405
VERSION BF063405.1 GI:10822315
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 489.

FEATURES
Source

1..565
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3323147"
/clone_lib="NCI-CCAP.Col16"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CCAP.Col16 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 152 a 138 c 115 g 160 t
ORIGIN

Query Match 8.8%; Score 557; DB 11; Length 565;
Best Local Similarity 100.0%; Pred. No. 1.3e-119;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5541 tctgcagatgatagcgaagcaatcctgtcctctgtcgtcttctctgtggtggtatactatgt 5600
Db 565 TCTGTCAGATGATGAGCGAAAGCATCTCTCTCTGCTCTGCTCTTCTGTGGGGCTATACATATGT 506
QY 5601 tctcagaacattgctgcccagaaattgtgaacggggaagaggttgagcttgcatcttt 5660
Db 505 TCTCAGAACATTTGCTGCCAGGAATTTGTAACGGGGAAGAGTTGGAGCTTGGCATTTT 446
QY 5661 cagcacttcactgtggagcggagctgctgcatcttctctaaatacagagaatgccgagtg 5720
Db 445 CAGGCATCTCACCTGTGAGCGGGAGTCTGCATTTTCTTAAATAACAGAGAATGCCGAGTG 386
QY 5721 gtcctgttggaaggttaaacgagagctgtgctctatccagctccttacttgatgaatat 5780
Db 385 GTCTGTGTTGAAGGTAAGGCAGAGGCTGTGCCTATCCAGCTCTCTTACTTGGATGAATAT 326
QY 5781 ggagaacagaccctggcctgaagaggggcaaccccttcatcttctcgtgagcggat 5840
Db 325 GGAGAACAACACCTTGGCCTGGAAGAGGGGCAACCCCTTCAATTATCTGTCGAGCGGAT 266
QY 5841 cggaactcctcattgttctgtgcaacaacactgcattatagaagagattgctaggagccaa 5900
Db 265 CGGAAGCTCCATTTGGTCTGCGCAACAACACTGCATATTAGAGAAGATGCTTAGGAGCCAA 206
QY 5901 gagactaatcagatgttatttgattcaactgagcagtgcttactgtgagctcccaactcgtc 5960
Db 205 GAGACTAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCGCT 146
QY 5961 caagacaatacaaaaatgacacagtagtaaaagcgtgattcaaaaattatggaactttct 6020
Db 145 CAAGACAATAACAAAATGACACAGTAGTAAGAGCGTGTATTCAAAATTTATGGAATACTTCT 86
QY 6021 gagggctgggaagatttgagggtcttttgcctccatgctcaggttcacttaccatcaata 6080
|||||


```

Db 85 GAGGCTGGGAAAGTAATTGGAGGCTCTTTTGCTCCATGTCAGGTTCACATTACATCAATA 26
Qy 6081 aaatatttcttaattgga 6097
|||||
Db 25 AAATATTCTTAATGGA 9

RESULT 7
BE589438 606 bp mRNA EST 28-AUG-2000
DEFINITION 195602 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE589438
VERSION BE589438.1 GI:9842477
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 606)
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
Mapping of Expressed Sequence Tags from a normalized bovine mammary
gland cDNA library
Unpublished (2000)
Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGCAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGAG
Plate: 120 row: N column: 11
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
Location/Qualifiers
1..606
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."
BASE COUNT 144 a 146 c 113 g 203 t
ORIGIN

Query Match 8.1%; Score 512.2; DB 10; Length 606;
Best Local Similarity 90.4%; Pred. No. 4.2e-109;
Matches 547; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 4761 aaagcattaatgcagttgcagttgcacagagagattacctgtctctcaggtccgtgatacag 4820
|||||
Db 1 AAGACATTAATGCAAGTTTGCCATTGCACAAAGGATTACCTGTTCTCAGGTCCGTGATACAG 60

Qy 4821 aaacatctgtctgtctctatcagttgtcttccttaacataaaacagagatacacca 4880
|||||
Db 61 AACATCTGATTCGCTTCATCATAGTTGTTCTTCCTAACCTGCAATCAGAGCTACACCT 120

Qy 4881 tgcctctgtctatagatctgtttcattgtttgtgtgggtgctgtgttagcattccccatcc 4940
|||||
Db 121 TGCCTCTATCCATAGATCATTTTCATGTTTGTGGGTACTGTGTAGCATTCCTCCGTCC 180

Qy 4941 ttgtattggatgaccctgtgtatctgcagcctctctctcagttagttcttctctataaccac 5000
|||||
Db 181 TTATACTGGGATGATGCTGTGTGATGCTGACGCTTCACCGTTAGCTTCTCGTATACACC 240

```

```

Qy 5001 ctttatctctccatttgatcacccatggcacacatgcttctteagatactactacagtagac 5060
|||||
Db 241 CTTTATCTCTTCCATTGATCACCATGGCACACATGCTTCAGATACTCTTTACCATAGAC 300

Qy 5061 acagccctaccccttgctcaggttcaagaagacagtgaaagaggtcattccgcattctct 5120
|||||
Db 301 ACAGACCTCCCTCTGCACAGATACAGGAAGAGTGAAGAGGCTCGTCTGCATCTTCT 360

Qy 5121 tcttttcagaaaatttctcaataacaaatggctccattgggtgtgatattctctggctgg 5180
|||||
Db 361 TTCCTGGCAGAAAGTTCTTCAGTATACAAAGTGGCTGCATGGGTGTGGTATTCTCTGCTGG 420

Qy 5181 tatttctgggtctcactgaagaatggcattccacccttatcttcgtgtgtgctgattgttt 5240
|||||
Db 421 TATTATGGGTCTCCTTGGAAGAAATGCAATCCCGCATCTCTCGCTGTGCTGCATTGTTT 480

Qy 5241 ttccactatttactgggttaactccgctcaggaactgcataccattctgcagaagga 5300
|||||
Db 481 TTCCACTATTTACTTGGAGTAACCTCCACCTGAAGAACTGTTTACCAATGCTGCAGAAGGA 540

Qy 5301 ggtacagtgactgtgtgctatctctcttaccacaaaattgttctcgtctctccag 5360
|||||
Db 541 GAATACAAATGTAATCTGTAGCTATTATTATCTTACCCACAAATTTGTTCTCTGCTTTTCCAG 600

Qy 5361 gaata 5365
|||||
Db 601 AAATA 605

```

RESULT 8

BG219270/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..624

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Athersys RAGE Library"

/cell_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression Nat. Biotechnol. 19 (5), 440 (2001) In press Contact: Scott J. Cain Athersys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scain@athersys.com High quality sequence stop: 624."

BASE COUNT 129 a 151 c 136 g 206 t

ORIGIN

<hr/>					
Query Match 7.8% Score 479..8; DB 11; Length 624;					
Best Local Similarity 98.2%; Pred. No. 1.6e-101;					
Matches 484; Conservative 0; Mismatches 9; Indels 0; Gaps 0;					
<hr/>					
Qy	3723	attactcatataagaaaaagcagaacgaaagaaaagctgaagctgctaggctacat	3782		
Db	624	ATTACTCATGATTAAGAAAAAGCAGAACGAAAAGAAAGCTGNAGCTGTAGGTTCAT	565		
Qy	3783	cgcagaagatcatggctcgagatgtctgcttcagaaaaacttcattgaacctcataaa	3842		
Db	564	CGCCAGGAATCATGGCTCAGATGCTGCCTTACAGAAAAAATTCATTGAACATCAATA	505		
Qy	3843	ctcatgtatgacaatacatcagaaaatccctggggaagaagtccattatggaggaagag	3902		
Db	504	CTCATGTATGACAATACATCAGAAATGCCCTGGGAAAGAAGATTCATTTATGGAGGAAG	445		
Qy	3903	agacccccagtcagtgactactctagaaatgcttttgggtcttaaacgggtgccatct	3962		
Db	444	AGCACCCCAGCAGTCAGTGACTACTCTAGAATTCCTTTGGGCTCTAAACGGGGTCCAATCT	385		
Qy	3963	gttactgaaaagagggtgctgtagcgtgcattcttgcacaagaagAACAGGAGTgaaaata	4022		
Db	384	GTPACTGAAAAGAGGAGTGTGACGTGCATCCTTTGCCCAAGAAGAACAGGAGGTGAAAATA	325		
Qy	4023	gaaaataatgccatgggtattatcggcgctgtgccagaaatctactgccttaaccacgac	4082		
Db	324	GAATAATGCCATGGTATTATCGGCCGTGTCTCCAGAAATCTACTGCTTAAACCCAGCAC	265		
Qy	4083	aggggaaacccatagaactctcaggagaagccctagaccacttttcattgatccagac	4142		
Db	264	AGGGGAANAACCCATAGAACTCTCAGSAGAGCCCTAGACCACATTTTCATGGATCCAGAC	205		
Qy	4143	tggccatatgaaacttatcacagaagctgtggtcatgtaatgcacgagtgctggcag	4202		
Db	204	TTGCCATATGCACTTATACAGGAGCTGTGGTCATGTATATGCACGACGAGTGTCTGGCAG	145		
Qy	4203	aagtattttgaag	4215		
Db	144	AACCTCTCCGGAG	132		
<hr/>					
RESULT	9				
AI192195/c					
LOCUS	AI192195	481 bp	mRNA	EST	28-OCT-1998
DEFINITION	qc92e08.x1 Soares.pregnant.uterus.NbHPU Homo sapiens cDNA clone IMAGE:1721702 3' similar to TR:O15057 O15057 KIA0349 ; , mRNA sequence.				
ACCESSION	AI192195				
VERSION	AI192195.1	GI:3743404			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov This clone is available royalty-free through LLNL : contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 563 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 351. Location/Qualifiers				
FEATURES	1 . .481				
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1721702" /clone_lib="Soares_pregnant_uterus_NbHPU" /sex="female"				

```

/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AATCGAAGAAATTCGGCGCCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Ratima Bonaldo."

BASE COUNT      125 a      96 c      86 g      174 t
ORIGIN

Query Match      7.6%; Score 477.8; DB 10; Length 481;
Best Local Similarity 99.6%; Pred. No. 4.6e-101;
Matches 479; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2778 cttcagattggtgcatttaattggtaccacaataagttcttgctacttggtacttcagagg 2837
Db 481 CTTcagATTGGTgcATCTTAAATGATGATCCAAATAGATTCCTGTCTACTGGTACTTCAGAGG 422

Qy 2838 tatgaacttgcgagggttttaacaagaccatatctacaaaagaccaggatttgatttaa 2897
Db 421 TATGAACTTGCGGAGGCTTTTAAACAAGACCATATCTACAAAAGCCAGGATTTGATTAAA 362

Qy 2898 caataataacacataagagaataatcttcaggctcctcatctatttggtggagcgt 2957
Db 361 CAATATAATACACTAATAGAGAATAATGCTTCAGGCTCCTCATCTATATGTGGGTGAGCGT 302

Qy 2958 tatgtacctggagtgggaaatgtgaccaaagaagaggtgcacaatgagagaataatcatcac 3017
Db 301 TATGTACTGGAGTGGGAATGTGACCAAGAAGAGGTCACAATGAGAGAATCATTCAC 242

Qy 3018 ttgctttgcatggaacccatgcccacacagtgccatttgccaaaaatttacctgagaatgaa 3077
Db 241 TTGCTTTGCATTGAACCCATGCCACACAGTGCCATTGCGCAAAATTTACCTTGAGATGAA 182

Qy 3078 aataataaactggcttagagaatgtcataaacaagaatggccacatttaagaaaccagggt 3137
Db 181 AATAATGAACCTGGCTTAGAGAATGTCTATAACAAAGTGCCACATTTTAAAGAAACCAAGT 122

Qy 3138 gtatcaggccatggagtttatgaactaaaagatgaactgaagacttcaatgatgtac 3197
Db 121 GTATCAGCCCATGAGTTTATGACTAAAGATGAATCAGTCAAGAGCTTCAATATGTAC 62

Qy 3198 ttttattactactccaaacccagcatagcaaggtcgaacatgtagaagaaagagaga 3257
Db 61 TTTTATCATTACTCCAAAACCCAGCATAGCAAGGCTGAACATATATGCAGAGAAAGGAGA 2

Qy 3258 a 3258
Db 1 A 1

RESULT 10
AW971391
LOCUS AW971391 583 bp mRNA EST 01-JUN-2000
DEFINITION EST383480 MAGE resequences, MAGL Homo sapiens cDNA, mRNA sequence.
ACCESSION AW971391
VERSION AW971391.1 GI:8161236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 583)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

```

JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 289

Seq primer: Forward.

Location/Qualifiers

FEATURES
source
1. .583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE reserences, MAGL"
/note="Vector: pBluescriptSKm"
171 a 126 c 137 g 149 t

Query Match 7.5%; Score 473.4; DB 10; Length 583;
Best Local Similarity 99.8%; Pred. No. 5.1e-100;
Matches 474; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5678 agccgagctgcatttctctaaatcagagaatccagtgctcgtgtaagataa 5737

Db 105 AGCCGAGGCTGCTTCTTAAATCAGAGATCCGAGTGGTCTGTTGAAGTAA 164

QY 5738 agccagagctgctctatccagctccttacttgatgtaataaggagaacagaccctgg 5797

Db 165 AGCCAGAGCTGTCCTATCCAGCTCCTTACTTGGATGATATGGAGAACAGACCCCTGG 224

QY 5798 cctgaagaggggcaacccctcattatctcgtgagcgatcgagaagctccattgggt 5857

Db 225 CTTGAAGAGGGGCAACCCCTTCATTATCTCGTGAGCGGTATCGGAAGCTCCATTGGT 284

QY 5858 ctggcaacaactgcattataagaagattgctagagagcaagagactaatcagatgtt 5917

Db 285 CTGGCAACAACACTGCTATTATAGAAGAGATGCTAGAGCCAGAGACTAATCAGATGTT 344

QY 5918 atttgattcaactggcagtgactgagctccaactctgcctcaagacatacacaaatg 5977

Db 345 ATTGTGATCACTGGCAGTACTGTGAGCTCCAACCTCGCTCAGACAAATCACAATG 404

QY 5978 agcagagtgaagagctgattcaaaattatggaaaacttctgagggctgggaaaagtat 6037

Db 405 ACACAGTAGTAAAGGCTGATTCAAAATTATGGAATAATTCTGAGGGCTGGGAAAGTAT 464

QY 6038 tggaggtctttgtccatgctccaggttcacttacatcaataaattatttctaattgga 6097

Db 465 TGGAGGGTCTTTTGTCCATGTCCAGGTTCCATCATCAATAAATAATTTCTTAATGGA 524

QY 6098 gtattgcttcaattagcaacataatgcttcacaggaagaaagagacatagatcaa 6152

Db 525 GTATTGCTTTCATTAGCAACATATGCTTCACAGGAAAGGACATAGATCCA 579

RESULT 11
AA401319
LOCUS AA401319 478 bp mRNA EST 09-NOV-1997
DEFINITION Zu63d04.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742663
5', mRNA sequence.
ACCESSION AA401319
VERSION AA401319.1 GI:2053683
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM

REFERENCE
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kriman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.

TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1003 Std Error: 0.00
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 458.
Location/Qualifiers

FEATURES
source
1. .478

/organism="Homo sapiens"
/db_xref="GDB:5929590"
/db_xref="taxon:9606"
/clone="IMAGE:742663"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaudo."

BASE COUNT 114 a 109 c 114 g 141 t
ORIGIN

Query Match 7.5%; Score 471.6; DB 10; Length 478;
Best Local Similarity 99.2%; Pred. No. 1.3e-99;
Matches 474; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5173 ctgctggtatttgggtctcactgaagaatggcatcaccccttatcttcgctgctg 5232

Db 1 CTGCTGTGATTGTGGGTCTCCTGGAAGATGGCATCACCCCTATCTTCCTGTGCTG 60

QY 5233 cattgttttccactatttacttgggttaactccgctgaggaactgcatacaattctg 5292

Db 61 CATTTGTTTTCACATATTTACTTGGGTAACTCCGCTCAGGAACCTGCATACTCTG 120

QY 5293 cagaagagagtagagtgacactctgtagctatctatttaccatacaaatgttctctgc 5352

Db 121 CAGAAGGAGAGTACAGTGCACCTCTAGCTATCTATCTTTACCTGCAAAATTTGTTCTCTGC 180

QY 5353 tcttccaggaatattgggactctgaagcccttgcctcagaggtggtgtcagatccctg 5412

Db 181 TCTTCCAGGAATATGGGATACTGTAAGCCCTTCTCCAGAGGTGGTGTGAGATCTCTG 240

QY 5413 ccttactaaactgtttgaagcaaaaaacacccgtgttcagggtaccctagaaaaaata 5472

Db 241 CCTTACTAAACTGTTGAAGCAAAAAACACCGTGGTCAGGTACCTAGAAAAAGAAATA 300

QY 5473 gttgatagagcttctgatgactatagctgctcctgaatcaagcttctcattcaggt 5532

Db 301 GTTTGATAGAGCTTCTCTGATGATAGTGTGCTCCTGCTCAAGTCTCTCATTTTCAGGT 360

QY 5533 gccacaggtctgcagatgatgagcgaaagcatcctgtcctctgctcttctcttgagggtc 5592

Db 361 GCCCAGCGTCTGCAGATGATGAGCGAAAAACAGTCTCTGCTCTGCTCTCTGCGGCTA 420

QY 5593 tactatgtcttcagaacacatttgcgtccaggaattgtgaacggggaagaggttgagc 5650

Db 421 TACTATGTTCTCAGAACATTTTGTCTCCAGGAAATTTGTGACGGGGAAGAGGTTGGAGC 478

```
RESULT 12
AW291190/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 522)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
source
Location/Qualifiers
1..522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2723895"
/clone_lib="NCI-CGAP_Sub4"
/clone_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The
NCI-CGAP_Sub4 library is a subtracted library derived from
the NCI-CGAP_Sub2 library which is a subtracted library
derived from the NCI-CGAP_Sub1 library, which is a
subtracted library derived from BI. BI constitutes a
mixture of 21 normalized or subtracted NCI-CGAP
libraries: NCI-CGAP_Co4, NCI-CGAP_Pr22, NCI-CGAP_Pr28,
NCI-CGAP_Co10, NCI-CGAP_Co16, NCI-CGAP_Kid5,
NCI-CGAP_Kid12, NCI-CGAP_Kid3, NCI-CGAP_Kid11,
NCI-CGAP_Lym2, NCI-CGAP_Br2, NCI-CGAP_Co8, NCI-CGAP_CLL1,
NCI-CGAP_Le12, NCI-CGAP_Brn23, NCI-CGAP_Lu5,
NCI-CGAP_Lu24, NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6,
NCI-CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI-CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE CloneIDs 1322376-1323911,
1456008-1456775, 1500552-1502855) NCI-CGAP_Kid5 pool 1 :
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs
1323912-1325831, 1471368-1472903, 1492104-1493255)
NCI-CGAP_Lu5 pool 1 : LLAM 3375-3582, 3851-3854 (IMAGE
CloneIDs 1414920-1417991, 1520904-1522439) NCI-CGAP_GC4
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
CloneIDs 1257096-1258631, 1459064-1470983, 1475592-1476743
) NCI-CGAP_Pr22 pool 1 : LLAM 2457-2459, 2758-2759,
3062-3068 (IMAGE CloneIDs 985608-986759, 1101192-1101959,
1217928-1220615) NCI-CGAP_Co10 pool 1 : LLAM 2644-2653,
2871-2872 (IMAGE CloneIDs 1057416-1061255, 1144584-1145351
) Subtraction was performed as previously described
[Bonaldo, Lennon & Soares (1996): Normalization and
Subtraction: Two Approaches To Facilitate Gene Discovery.
Genome Research 6, 791-806.]
TAG_LIB=NCI-CGAP_Co4
TAG_TISSUE=colon
TAG_SEQ=CTTCG"
137 a 125 c 100 g 160 t
BASE COUNT
```

ORIGIN

```
Query Match 7.2%; Score 454; DB 10; Length 522;
Best Local Similarity 98.0%; Pred. No. 1.7e-95;
Matches 492; Conservative 0; Mismatches 5; Indels 5; Gaps 3;

Qy 5596 tatgtctcagaacattgtgctccaggaattgtgaacggggaagaggttggagcttgca 5655
|||||
Db 513 TATGTTCTCAGAACATTTGCTGCCAGGAAA-TGTGAACGGGGAAGAGTTGGAG--TTCA 458
|||||
Qy 5656 ttttcacgcacttcactgtgagccggagtgctgcatcttctctaaataacagagaatgcc 5715
|||||
Db 457 TTTTTCACGCACCTTCACTGTGAGCGGAGTCTGCATTTTCTCTAAAAATCAGAGAATGCC 398
|||||
Qy 5716 gagtgctcctgggtgaaggtaaaagccagaggtgtgctctatccagctctcttaattggatg 5775
|||||
Db 397 GAGTGTCTCTGGTTGAAGGTAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTGGATG 338
|||||
Qy 5776 aataggagaacacagaccctggcctgaagaggggcaaccccttctattctcgtgagc 5835
|||||
Db 337 AATATGGAGAAACACACCCCTGCCCTGAAGAGGGGCAACCCCTTCATTTATCTCTGAGC 278
|||||
Qy 5836 ggtatcggaagctccatttgggtctgtgcaaacacactgcattatagaagagattgctagga 5895
|||||
Db 277 GGTATCGGAAGCTCCATTTGGTCTGSCAACACACTGCAATATAGAAGAGATTGCTAGGA 218
|||||
Qy 5896 gccaaagactaaatcagatgttatttggattcaactggcagttactgtgagctccaaactc 5955
|||||
Db 217 GCCAAGACACTAATCAGATGTTATTGGATTCAACTGGCAGTTACTGTGAGCTCCAATC 158
|||||
Qy 5956 tgcctcagaacaatcacaaatgacacagctagtaaggctgattcaaaattatggaaaac 6015
|||||
Db 157 TGCCTCAGACAATCACAATGACGACAGTAGTAAGGCTGATTCAAAATATTGGAAC 98
|||||
Qy 6016 ttctgagggctgggaagattgagggctcttttgcctccatgtccaggttcacttacat 6075
|||||
Db 97 TTTCTGAGGGCTGGGAAAGTATTGGA-GGCTTTTGTCCATGTCCAGGTTCACTTACAT 39
|||||
Qy 6076 caataaataattcttaattgga 6097
|||||
Db 38 CAATAAAATATTCTTAATGGA 17
|||||

RESULT 13
BG382648
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 522)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.
EST.
BG382648.1 GI:13307120
EST.
p1g.
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 522)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.
EST discovery in swine
Design and use of two pooled tissue normalized cDNA libraries for
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
```

FEATURES		TITLE	
source		JOURNAL	
BACKWARD: GTTTTCCAGTCACGACG		COMMENT	
Plate: 4 row: H column: 19		Design and use of two pooled tissue normalized cDNA libraries for	
Seq primer: ATTAGTGACACTATAG.		EST discovery in swine	
Location/Qualifiers		Unpublished (2000)	
1. 522		Contact: Smith TPL	
/organism="Sus scrofa"		USDA, ARS, US Meat Animal Research Center	
/db_xref="taxon:9823"		PO Box 166, Clay Center, NE 68933-0166, USA	
/clone_lib="MARC lPIG"		Tel: 402 762 4366	
/tissue_type="pooled"		Fax: 402 762 4390	
/lab_host="DH10B"		Email: smith@email.marc.usda.gov	
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;		Single pass sequencing. Bases called and alt_trimmed with phred	
Library made from pooled tissue from day 11, 13, 15, 20,		v0.980904.e. Vector identified by cross_match with the -minscore 18	
and 30 embryos."		and -minmatch 12 options.	
PCR primers			
FORWARD: AGAAACAGCTATGACCAT			
BACKWARD: GTTTTCCAGTCACGACG			
Plate: 4 row: B column: 23			
Seq primer: ATTAGTGACACTATAG.			
Location/Qualifiers			
1. 520			
/organism="Sus scrofa"			
/db_xref="taxon:9823"			
/clone_lib="MARC lPIG"			
/tissue_type="pooled"			
/lab_host="DH10B"			
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;			
Library made from pooled tissue from day 11, 13, 15, 20,			
and 30 embryos."			
BASE COUNT		121 a 126 c 99 g 174 t	
ORIGIN			
Query Match		7.1%; Score 449; DB 11; Length 522;	
Best Local Similarity		91.4%; Pred. No. 2.6e-94;	
Matches 476; Conservative 0; Mismatches 45; Indels 0; Gaps 0;			
QY	4919	tgctgtgttagcattcccatccctctgtattgggatgacccctgttgatctgcagccctcttc	4978
Db	1	TGCTGTGTAGCCTTCCCATCCTTGTATTGGGATGATGCTGTGTGATCTGCAGCCCTCCTC	60
QY	4979	agttagttctctataaccaccccttctctccattgtatccattgtacacacacacatgct	5038
Db	61	AGTTAGCTCTTCCCTATAACACACCTTTATCTCTTCCATTTGATCACCATGGCAGATGT	120
QY	5039	tcagatactactacagtagacagcagcctaccctctgctcaggttcgaagacagctga	5098
Db	121	TCAGATACCTTCTTACCATAGACACAGGCCTTCCCTTGCACAGACTCAAGAAGAGAGTGA	180
QY	5099	agaggtcattccgcacatctctctcttcttgagaaattctcaataatacaaatggctccat	5158
Db	181	AGAGGCTCATCTGCACTCTTCTTCTTGGCAGAAAGTTCTCAATATACAAATGGCTAGCT	240
QY	5159	tgggtgtgatctctctgctggttattgtggtctcactgaagaatggcatcacccctta	5218
Db	241	TGGGTGTGATTTCCTGGCTGGTATTATTTATGGGTCTCATTTGAAGATGGCATATCCCTTA	300
QY	5219	tctcgtctgctgctcattgttttccactatttacttgggtgaactccgcctggagaaact	5278
Db	301	CCITTCGCTGTGCTGCACTTGTCTTCTTCCATTTATTTACTTGGAAATACTCCACCTGAAGATCT	360
QY	5279	gcataccaaattctgcagaagagagtagcagtgactctgtagctatctattcttacctac	5338
Db	361	GTTTACCAATCTGCAGAGAGGAGTACAGTGCACCTCTGTAGCTATTATTATCTTTACCCAC	420
QY	5339	aaattgttctcgtctctctccaggaataatgggatactgtaaggcccttgcctcagaggtg	5398
Db	421	AAATTTGTTCTGCTTTTCCAGAAATATTGGGATACTATAAGGCCCTCTCCAGAGGTG	480
QY	5399	gtgtgcagatctcgtccttactaaactgtttgaagcaaaaa	5438
Db	481	GTGTGCAGACCCCTGCCTTACTAAACTGTTTGAAGCAAAAA	520
RESULT		15	

FEATURES		TITLE	
source		JOURNAL	
BACKWARD: GTTTTCCAGTCACGACG		COMMENT	
Plate: 4 row: H column: 19		Design and use of two pooled tissue normalized cDNA libraries for	
Seq primer: ATTAGTGACACTATAG.		EST discovery in swine	
Location/Qualifiers		Unpublished (2000)	
1. 522		Contact: Smith TPL	
/organism="Sus scrofa"		USDA, ARS, US Meat Animal Research Center	
/db_xref="taxon:9823"		PO Box 166, Clay Center, NE 68933-0166, USA	
/clone_lib="MARC lPIG"		Tel: 402 762 4366	
/tissue_type="pooled"		Fax: 402 762 4390	
/lab_host="DH10B"		Email: smith@email.marc.usda.gov	
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;		Single pass sequencing. Bases called and alt_trimmed with phred	
Library made from pooled tissue from day 11, 13, 15, 20,		v0.980904.e. Vector identified by cross_match with the -minscore 18	
and 30 embryos."		and -minmatch 12 options.	
PCR primers			
FORWARD: AGAAACAGCTATGACCAT			
BACKWARD: GTTTTCCAGTCACGACG			
Plate: 4 row: B column: 23			
Seq primer: ATTAGTGACACTATAG.			
Location/Qualifiers			
1. 520			
/organism="Sus scrofa"			
/db_xref="taxon:9823"			
/clone_lib="MARC lPIG"			
/tissue_type="pooled"			
/lab_host="DH10B"			
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;			
Library made from pooled tissue from day 11, 13, 15, 20,			
and 30 embryos."			
BASE COUNT		122 a 126 c 100 g 174 t	
ORIGIN			
Query Match		7.1%; Score 449; DB 11; Length 522;	
Best Local Similarity		91.4%; Pred. No. 2.6e-94;	
Matches 476; Conservative 0; Mismatches 45; Indels 0; Gaps 0;			
QY	4919	tgctgtgttagcattcccatccctctgtattgggatgacccctgttgatctgcagccctcttc	4978
Db	1	TGCTGTGTAGCCTTCCCATCCTTGTATTGGGATGATGCTGTGTGATCTGCAGCCCTCCTC	60
QY	4979	agttagttctctataaccaccccttctctccattgtatccattgtacacacacacatgct	5038
Db	61	AGTTAGCTCTTCCCTATAACACACCTTTATCTCTTCCATTTGATCACCATGGCAGATGT	120
QY	5039	tcagatactactacagtagacagcagcctaccctctgctcaggttcgaagaagacagtg	5098
Db	121	TCAGATACCTTCTTACCATAGACACAGGCCTTCCCTTGCACAGACTCAAGAAGAGAGTGA	180
QY	5099	agaggtcattccgcacatctctctcttcttgagaaattctcaataatacaaatggctccat	5158
Db	181	AGAGGCTCATCTGCACTCTTCTTCTTGGCAGAAAGTTCTCAATATACAAATGGCTAGCT	240
QY	5159	tgggtgtgatctctctgctggttattgtggtctcactgaagaatggcatcacccctta	5218
Db	241	TGGGTGTGATTTCCTGGCTGGTATTATTTATGGGTCTCATTTGAAGATGGCATATCCCTTA	300
QY	5219	tctcgtctgctgctcattgttttccactatttacttgggtgaactccgcctggagaaact	5278
Db	301	CCITTCGCTGTGCTGCACTTGTCTTCTTCCATTTATTTACTTGGAAATACTCCACCTGAAGATCT	360
QY	5279	gcataccaaattctgcagaagagagtagcagtgactctgtagctatctattcttacctac	5338
Db	361	GTTTACCAATCTGCAGAGAGGAGTACAGTGCACCTCTGTAGCTATTATTATCTTTACCCAC	420
QY	5339	aaattgttctcgtctctccaggaataatgggatactgtaaggcccttgcctcagaggtg	5398
Db	421	AAATTTGTTCTGCTTTTCCAGAAATATTGGGATACTATAAGGCCCTCTCCAGAGGTG	480
QY	5399	gtgtgcagatctcgtccttactaaactgtttgaagcaaaaa	5439
Db	481	GTGTGCAGACCCCTGCCTTACTAAACTGTTTGAAGCAAAAA	521
RESULT		14	

FEATURES		TITLE	
source		JOURNAL	
BACKWARD: GTTTTCCAGTCACGACG		COMMENT	
Plate: 4 row: H column: 19		Design and use of two pooled tissue normalized cDNA libraries for	
Seq primer: ATTAGTGACACTATAG.		EST discovery in swine	
Location/Qualifiers		Unpublished (2000)	
1. 522		Contact: Smith TPL	
/organism="Sus scrofa"		USDA, ARS, US Meat Animal Research Center	
/db_xref="taxon:9823"		PO Box 166, Clay Center, NE 68933-0166, USA	
/clone_lib="MARC lPIG"		Tel: 402 762 4366	
/tissue_type="pooled"		Fax: 402 762 4390	
/lab_host="DH10B"		Email: smith@email.marc.usda.gov	
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;		Single pass sequencing. Bases called and alt_trimmed with phred	
Library made from pooled tissue from day 11, 13, 15, 20,		v0.980904.e. Vector identified by cross_match with the -minscore 18	
and 30 embryos."		and -minmatch 12 options.	
PCR primers			
FORWARD: AGAAACAGCTATGACCAT			
BACKWARD: GTTTTCCAGTCACGACG			
Plate: 4 row: B column: 23			
Seq primer: ATTAGTGACACTATAG.			
Location/Qualifiers			
1. 520			
/organism="Sus scrofa"			
/db_xref="taxon:9823"			
/clone_lib="MARC lPIG"			
/tissue_type="pooled"			
/lab_host="DH10B"			
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;			
Library made from pooled tissue from day 11, 13, 15, 20,			
and 30 embryos."			
BASE COUNT		122 a 126 c 100 g 174 t	
ORIGIN			
Query Match		7.1%; Score 449; DB 11; Length 522;	
Best Local Similarity		91.4%; Pred. No. 2.6e-94;	
Matches 476; Conservative 0; Mismatches 45; Indels 0; Gaps 0;			
QY	4919	tgctgtgttagcattcccatccctctgtattgggatgacccctgttgatctgcagccctcttc	4978
Db	1	TGCTGTGTAGCCTTCCCATCCTTGTATTGGGATGATGCTGTGTGATCTGCAGCCCTCCTC	60
QY	4979	agttagttctctataaccaccccttctctccattgtatccattgtacacacacacatgct	5038
Db	61	AGTTAGCTCTTCCCTATAACACACCTTTATCTCTTCCATTTGATCACCATGGCAGATGT	120
QY	5039	tcagatactactacagtagacagcagcctaccctctgctcaggttcgaagaagacagtg	5098
Db	121	TCAGATACCTTCTTACCATAGACACAGGCCTTCCCTTGCACAGACTCAAGAAGAGAGTGA	180
QY	5099	agaggtcattccgcacatctctctcttcttgagaaattctcaataatacaaatggctccat	5158
Db	181	AGAGGCTCATCTGCACTCTTCTTCTTGGCAGAAAGTTCTCAATATACAAATGGCTAGCT	240
QY	5159	tgggtgtgatctctctgctggttattgtggtctcactgaagaatggcatcacccctta	5218
Db	241	TGGGTGTGATTTCCTGGCTGGTATTATTTATGGGTCTCATTTGAAGATGGCATATCCCTTA	300
QY	5219	tctcgtctgctgctcattgttttccactatttacttgggtgaactccgcctggagaaact	5278
Db	301	CCITTCGCTGTGCTGCACTTGTCTTCTTCCATTTATTTACTTGGAAATACTCCACCTGAAGATCT	360
QY	5279	gcataccaaattctgcagaagagagtagcagtgactctgtagctatctattcttacctac	5338
Db	361	GTTTACCAATCTGCAGAGAGGAGTACAGTGCACCTCTGTAGCTATTATTATCTTTACCCAC	420
QY	5339	aaattgttctcgtctctccaggaataatgggatactgtaaggcccttgcctcagaggtg	5398
Db	421	AAATTTGTTCTGCTTTTCCAGAAATATTGGGATACTATAAGGCCCTCTCCAGAGGTG	480
QY	5399	gtgtgcagatctcgtccttactaaactgtttgaagcaaaaa	5439
Db	481	GTGTGCAGACCCCTGCCTTACTAAACTGTTTGAAGCAAAAA	521
RESULT		14	

FEATURES		TITLE	
source		JOURNAL	
BACKWARD: GTTTTCCAGTCACGACG		COMMENT	
Plate: 4 row: H column: 19		Design and use of two pooled tissue normalized cDNA libraries for	
Seq primer: ATTAGTGACACTATAG.		EST discovery in swine	
Location/Qualifiers		Unpublished (2000)	
1. 522		Contact: Smith TPL	
/organism="Sus scrofa"		USDA, ARS, US Meat Animal Research Center	
/db_xref="taxon:9823"		PO Box 166, Clay Center, NE 68933-0166, USA	
/clone_lib="MARC lPIG"		Tel: 402 762 4366	
/tissue_type="pooled"		Fax: 402 762 4390	
/lab_host="DH10B"		Email: smith@email.marc.usda.gov	
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;		Single pass sequencing. Bases called and alt_trimmed with phred	
Library made from pooled tissue from day 11, 13, 15, 20,		v0.980904.e. Vector identified by cross_match with the -minscore 18	
and 30 embryos."		and -minmatch 12 options.	
PCR primers			
FORWARD: AGAAACAGCTATGACCAT			
BACKWARD: GTTTTCCAGTCACGACG			
Plate: 4 row: B column: 23			
Seq primer: ATTAGTGACACTATAG.			
Location/Qualifiers			
1. 520			
/organism="Sus scrofa"			
/db_xref="taxon:9823"			
/clone_lib="MARC lPIG"			
/tissue_type="pooled"			
/lab_host="DH10B"			
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;			

```

BE550873/c
LOCUS       BE550873             486 bp      mRNA           EST           06-SEP-2000
DEFINITION  UI-M-BH3-ata-a-07-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
ACCESSION   BE550873
VERSION     BE550873.1  GI:9976697
KEYWORDS    house mouse.
SOURCE      Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ORGANISM    1 (bases 1 to 486)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
REFERENCE   Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            9704477
            Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-ata-a-07-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pR73D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately derived from a mixture of individually tagged
                     normalized libraries from ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland, striatum,
                     hippocampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows: PCR amplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (Life Technologies) to generate the
                     NIH_BMAP_M_S4 library. This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)."
BASE COUNT      129 a      117 c      95 g      145 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Search completed: March 1, 2002, 14:46:36
Job time: 8041 sec

Query Match 6.5%; Score 410.8; DB 10; Length 486;
Best Local Similarity 90.3%; Pred. No. 2.3e-85;
Matches 439; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 13:51:46 ; Search time 261.48 Seconds
(without alignments)
20682.299 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattcgcacgagg.....aatttgatttggtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 42862619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4435.4	70.3	6395	20	AAV99308
2	4435.4	70.3	6395	22	AAV99308
3	2546.8	40.4	2550	22	AAH14878
4	996.2	15.8	1001	20	AAH03300
5	996.2	15.8	1001	22	AAH03300
6	813.6	12.9	3327	20	AAH35730
7	813.6	12.9	3327	22	AAH35730
8	656.8	10.4	733	21	AAH02411
9	647	10.3	712	22	AAH07621
10	638.6	10.1	756	21	AAH02327
11	246.8	3.9	807	20	AAH39891

12	241.6	3.8	264	21	AAA00233	Human colon cancer
13	191.6	3.0	800	20	AAV39892	Gastric cancer ass
14	159	2.5	455	20	AAV89674	EST clone DA490.
15	118	1.9	936	22	AAF58252	Oligonucleotide D1
16	118	1.9	936	22	AAF58254	Oligonucleotide D1
17	118	1.9	936	22	AAF58257	Oligonucleotide D1
18	118	1.9	936	22	AAF58259	Oligonucleotide D2
19	118	1.9	936	22	AAF58262	Oligonucleotide D2
20	118	1.9	938	22	AAF58255	Oligonucleotide D1
21	116.6	1.8	308	16	AAT21876	Human gene signatu
22	116.4	1.8	936	22	AAF58252	Oligonucleotide D1
23	116.4	1.8	936	22	AAF58254	Oligonucleotide D1
24	116.4	1.8	936	22	AAF58257	Oligonucleotide D1
25	116.4	1.8	936	22	AAF58259	Oligonucleotide D2
26	116.4	1.8	936	22	AAF58262	Oligonucleotide D2
27	116.4	1.8	938	22	AAF58255	Oligonucleotide D1
28	105.6	1.7	366	21	AAV00938	Human secreted pro
29	105.6	1.7	456	21	AAV00940	Human secreted pro
30	105.6	1.7	1461	21	AAV78096	Human cancer assoc
31	103.2	1.6	605	21	AAV00939	Human secreted pro
32	100.8	1.6	602	21	AAV10733	Human secreted pro
33	100.8	1.6	692	21	AAV10734	Human secreted pro
34	100.8	1.6	842	21	AAV10872	Human secreted pro
35	54.8	0.9	2793	21	AAV70146	Plasmodium falcipa
36	52.4	0.8	244	22	AAF58238	Oligonucleotide D1
37	51.6	0.8	244	22	AAF58238	Oligonucleotide D1
38	49.2	0.8	1727	22	AAH17089	Human CDNA sequenc
39	48.6	0.8	4590	22	AAH24065	Yeast AOP9604-asso
40	46.2	0.7	581	22	AAV09232	Human CDNA clone (
41	45.2	0.7	11722	19	AAV34455	Human MHC class I
42	45	0.7	143068	21	AAF21105	Human low adenosin
43	45	0.7	143068	21	AAF21272	Human low adenosin
44	45	0.7	143068	21	AAV34983	Human adenosine re
45	45	0.7	143068	21	AAV35150	Human adenosine re

ALIGNMENTS

RESULT 1						
AAV99308						
ID	AAV99308	standard;	cdna;	6395	BP.	
XX						
AC	AAV99308;					
XX						
DT	25-MAR-1999	(first entry)				
XX						
DE	cdna encoding mouse a ubiquitin-protein ligase, Ubr1.					
XX						
KW	Ubiquitin-protein ligase; Ubr1; mouse; ubiquitinylation; degradation;					
KW	N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.					
XX						
OS	Mus sp.					
XX						
PH	Key					
FT	CDS					
FT						
FT						
XX						
XX						
PN	US5861312-A.					
XX						
PD	19-JAN-1999.					
XX						
PF	02-DEC-1997;					
XX						
PR	02-DEC-1997;					
XX						
PA	(CALY) CALIFORNIA INST OF TECHNOLOGY.					
XX						
PI	Kwon YT, Varshavsky A;					
XX						
DR	WPI; 1999-130395/11.					
DR	P-PSDB; AAW84351.					

XX	Mouse and human Ubr1 cDNA - useful for producing recombinant Ubr1 polypeptides
PT	Claim 1; Columns 15-28; 18pp; English.
XX	The present sequence encodes a ubiquitin-protein ligase called Ubr1. The Ubr1 enzymes are involved in protein ubiquitinylation and ultimate degradation through the N-end rule pathway and have been linked to stress-related muscle wasting. Recombinant Ubr1 polypeptides can be used to screen for inhibitors of muscle wasting when this is associated with the N-end rule pathway.
XX	Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;
XX	Query Match 70.3%; Score 4435.4; DB 20; Length 6395; Best Local Similarity 86.3%; Pred. No. 0; Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps
QY	590 ttcaggggccgtctaaagtgtctccctctctc-tccgaccgcgcacaggtttcgcgt 648
DB	8 ttcaggggccgtctaaagtgtctccctctcgcgtcgcgccgcgcacaggtttcgcgt 67
QY	649 tgcctctgcccggggtcggaactcgaggcgtcagtttccctcaagatggcgacagg 708
DB	68 agctgcggccggggtcggaactcgggcgtctcgtttcccttaagatggcgacagg 127
QY	709 agcgtggaagtactgagaggatggaatcagcgcggaggtaccgccagccctcagcgc 768
DB	128 agatggacgcgcgcgagaggatggagcgtcagcccgagcgtccctcgtgcccgacgcgc 187
QY	769 tggcatctgtgggatcagcaagtgtattttactgcgtttcttgcatactttggcac 828
DB	188 cggcatcgtgggtgagcagcaagtgtattctatactgcgtttcttatactattggcac 247
QY	829 aattgggcacaaatttactttgtcgaatggaccacagacttggaaaaagcaggagaaa 888
DB	248 aattagtcgcaaaattattttgtcgtgagatggaccagatttggaaaaagcagaaga 307
QY	889 gttcaaatgtcaatatcactccactggaaatggtactatttggagagatccagata 948
DB	308 ggtcacagatgtcaatactactcctttggagtggtactatttggagagatcccgata 367
QY	949 tttgcttagagaaattgaagcacagtgagcatttcagctttgtggaggggtttcmeta 1008
DB	368 tttgcttagagaaattaaaaacacagtgagcgttccagttgtgtggagaggtttcmeta 427
QY	1009 gtggagagacacctattctgtcaggagattgtcaattgtatccaaatgtgtactctgta 1068
DB	428 gtggagaaacaatctctgtagggattgtcaattgtatccaaatgtgtgtgtgtgta 487
QY	1069 tggactgtcttcagagacagtgttcataaaatcactgttcaagatgcatacttctactg 1128
DB	488 tggactgtcttcacaaagtgtgtcataaaacactgttcaagatgcatacttctactg 547
QY	1129 gagggggttctgtgactgtgtgagacagagcgcatggaaactggcccttttggtaaa 1188
DB	548 gagggggttctgtgactgtgtgagacagagcggtgaaactggcccttttgggtgg 607
QY	1189 atcatgaacctggaagcaggttactataaaagagaattcacgctgtccgttgaatgaag 1248
DB	608 atcacgctgtggaagcaggttactataaaagagagcttacattgccattgaaatgaag 667
QY	1249 aggtaatgttcaaacccaggaaaaatttcttccgtcagtgataaaatatgtcgtagaaatga 1308
DB	668 aggtgatgtctcaagccaggagaatttcccttcggtgataaaatacatctgtgaaatga 727
QY	1309 ctatgtggaagagaaaaaagactcctcctgaactccagatgaaggagaaaaatgaaa 1368
DB	728 ctatatgggaagagaaaaagaaattgcctcctcgtactgcagatgaaggagaaaaatgaac 787
QY	1369 gatactattgtctcttccaatgatgaacaccattcatatgaccacgttcatacagcc 1428

Qy	4657	cctgagacacctg	cgctttca	ctatccagg	caattg	aaaaat	ctatttgg	gagatga	aaagaa	4716
Db	4088	cctggagcag	tgtgcgttca	ccattccag	ggaatcg	aaacctgt	tgga	gagatga	aaagaa	4147
Qy	4717	aacctcgtttg	agcaacttca	aaatag	cgacga	ataatg	tcttga	aagcatta	gcagt	4776
Db	4148	aacctcta	ttggagca	cttcaaa	tagacag	catag	cggtctg	gaagcgctaa	tcagct	4207
Qy	4777	ttagcttgca	cagagga	tta	ctgctc	caggtct	ctgata	cacaga	acatctgtt	4836
Db	4208	ttagcttgca	cagaggct	ac	ctgcct	caggtc	ctgata	cacaa	catctg	4267
Qy	4837	ttctata	ctgtt	cttctta	caata	taaaat	caga	agata	caccatgc	4896
Db	4268	tctgtcag	tattt	cttcta	acctg	caatcaga	aaaa	taca	ccagcctt	4327
Qy	4897	atcgtttc	atgtttt	gggtg	ctgtgt	ttagca	ttcc	atcctt	gtattgg	4956
Db	4328	atctctca	atgtct	tggtcg	gcag	ctctag	cggtcc	catc	ctgtattgg	4387
Qy	4957	ctgttgat	ctcag	ccttct	cagttag	ttctt	ctata	taacc	actttatct	5016
Db	4388	cgtg	atctgcag	cgctgc	caact	tagttct	caata	taacc	actctatct	4447
Qy	5017	tgtac	acatg	gcacac	atg	cttc	agata	tact	actta	5064
Db	4448	tgtac	acatg	gcgcac	atg	cttc	agata	tact	actta	4507
Qy	5065	gcctac	cccttgc	taag	tta	aga	agac	agctga	agagct	5124
Db	4508	ggccgc	gctgtg	aggg	tga	aggat	ag	ctgtgc	atctgt	4567
Qy	5125	ttaga	aaattt	ctca	ata	caga	tggt	ccattgg	gtgtat	5184
Db	4568	tgtg	gaagtgt	gcag	acac	agac	ggc	ctactgg	gtgcggt	4627
Qy	5185	tgtgg	gtctc	actga	gaat	ggcat	ccccc	tatctt	gcgtgtgc	5244
Db	4628	tgtg	gctctcc	ctgag	aac	ggcat	ccccc	tactt	ccgctgtg	4687
Qy	5245	actatt	ta	cttggg	ta	actcc	gc	ctga	gaactgc	5304
Db	4688	actatt	ta	cttgg	atag	ctc	gc	ctga	gaactgtt	4747
Qy	5305	acag	tc	actctgt	ag	ctat	ctt	ta	ccataaa	5364
Db	4748	ta	agtc	actctgt	ag	ctat	ctt	ta	ccataaa	4807
Qy	5365	attg	gata	ctgt	a	ggccct	gt	ccag	aggtg	5424
Db	4808	attg	gata	ctacc	ata	agg	cccttact	ac	aggtg	4867
Qy	5425	gttt	ga	gc	aaaaa	ac	ccg	tg	ctcag	5484
Db	4868	ctt	ga	gc	aaaaa	ag	tg	ctg	tg	4927
Qy	5485	ttc	ct	ga	ct	ga	ctc	ct	ga	5544
Db	4928	ttc	ct	ga	ga	ct	ga	ctc	ga	4987
Qy	5545	cag	at	ga	gc	ga	ag	ca	ctc	5604
Db	4988	cag	at	ga	gc	ga	ag	ca	ctc	5047
Qy	5605	aga	a	ctt	gt	gc	ag	ga	aaat	5664
Db	5048	aga	a	ctt	gt	gc	ag	ga	aaat	5107
Qy	5665	ca	ctt	ca	ct	gt	gg	ag	ct	5724
Db	5108	cg	ctt	ca	ct	gt	gg	ag	ct	5167
Qy	5725	tgg	t	ga	ag	gt	ta	aa	gc	5784

[illegible]


```
QY 2979 gtgaccaaagagaggtcacaatgagagaaatcattcacttgccttgcttgacccatg 3038
Db 181 gtgaccaaagagaggtcacaatgagagaaatcattcacttgccttgcttgacccatg 240
QY 3039 ccacacagtgcattcccaaaaattacctgagaaatgaaataatgaaactggcttagag 3098
Db 241 ccacacagtgcattcccaaaaattacctgagaaatgaaataatgaaactggcttagag 300
QY 3099 aatgtcataaaacaaagtgcaccacatttagaagaaacccgggtgtatcaggccatggagttat 3158
Db 301 aatgtcataaaacaaagtgcaccacatttagaagaaacccgggtgtatcaggccatggagttat 360
QY 3159 gaactaaaagatgaatcactgaaagactccaatgatctacttttattcattactccaaaacc 3218
Db 361 gaactaaaagatgaatcactgaaagactccaatgatctacttttattcattactccaaaacc 420
QY 3219 cagcatagcaaggtgaacatgacagaaagagaaagagaaacaaagaaacaaagatgaa 3278
Db 421 cagcatagcaaggtgaacatgacagaaagagaaagagaaacaaagaaacaaagatgaa 480
QY 3279 gaattgcgccaccaccacacctcctgaattcgtccctgcttccagcaaaagtgttaacctt 3338
Db 481 gaattgcgccaccaccacacctcctgaattcgtccctgcttccagcaaaagtgttaacctt 540
QY 3339 ctcaactgtgatcatgatgatcacattctcaggaccgtatttgagcggccaatagacaca 3398
Db 541 ctcaactgtgatcatgatgatcacattctcaggaccgtatttgagcggccaatagacaca 600
QY 3399 gattctaaacttggaccgaaggtgctccaaatggcttttcattcttgccttgcttgcctt 3458
Db 601 gattctaaacttggaccgaaggtgctccaaatggcttttcattcttgccttgcttgcctt 660
QY 3459 ttactagaagaagcaacagcttccaaaagctcctgaaagaagaatgaaacttgcctt 3518
Db 661 ttactagaagaagcaacagcttccaaaagctcctgaaagaagaatgaaacttgcctt 720
QY 3519 tatcataagctctcaagattggaaattcagccaatgaaataacaaatgcttttggaaaaa 3578
Db 721 tatcataagctctcaagattggaaattcagccaatgaaataacaaatgcttttggaaaaa 780
QY 3579 ctcaaggaatctccacagttagaagccgaagacatgatacgtggatacttcagatg 3638
Db 781 ctcaaggaatctccacagttagaagccgaagacatgatacgtggatacttcagatg 840
QY 3639 ttgacacagtgaaagcgttaagagaaaaatcttgttattgtagcaaccacatcagga 3698
Db 841 ttgacacagtgaaagcgttaagagaaaaatcttgttattgtagcaaccacatcagga 900
QY 3699 tcggaatctattaagaatgatgagattactcatgataaagaaaaagcagaacgaaaaaga 3758
Db 901 tcggaatctattaagaatgatgagattactcatgataaagaaaaagcagaacgaaaaaga 960
QY 3759 aaagctgaagctgctagctacatcccgagaagatcgtgc 3799
Db 961 aaagctgaagctgctagctacatcccgagaagatcgtgc 1001
```

RESULT 5

```
AAC86934
ID AAC86934 standard; cDNA; 1001 BP.
XX
XX AAC86934;
XX
XX 02-APR-2001 (first entry)
XX
XX Nucleotide sequence of a partial human Ubql protein.
XX
XX Ubql; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
KW Versinia enterocolitica; muscle wasting; infection; ss.
XX
XX Homo sapiens.
XX
```

```
PH Key Location/Qualifiers
FT CDS 1..999
FT /*tag= a
FT /product= "Ubql"
FT /note= "partial sequence"
XX
XX US6159732-A.
XX
XX 12-DEC-2000.
XX
XX 11-JAN-1999; 99US-0228317.
XX
XX 02-DEC-1997; 97US-0982956.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Kwon YT, Varshavsky A;
XX
XX WPI; 2001-090278/10.
XX
XX P-PSDB; AAB31163.
XX
XX Inhibiting the N-end rule pathway in mammalian cells for treating
XX infections and various diseases associated with muscle tissue wasting,
XX by inhibiting the expression of Ubql gene
XX
XX Claim 4; Column 27-30; 18pp; English.
XX
XX The present sequence encodes a partial Ubql enzyme. Ubql is an E3-type
XX protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
XX ligase. The enzyme is specific for destabilizing residues exposed at
XX the N-terminus of protein substrates. Inhibition of the expression of
XX Ubql gene in a cell results in inhibition of the N-end rule pathway.
XX The method is used for treatment of mammalian cells infected with an
XX intracellular pathogen, e.g. Lysteria monocytogenes or Versinia
XX enterocolitica. Inhibition of N-end rule pathway is also useful for
XX treating various diseases associated with wasting of muscle tissue and
XX infections.
XX
XX Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;
```

```
Query Match 15.8%; Score 996.2; DB 22; Length 1001;
Best Local Similarity 99.7%; Pred. No. 1.9e-261;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2799 atggtatcccaataagcttcttactgtacttcagaggtatgaaacttgcaggctttt 2858
Db 1 atggtatcccaataagcttcttactgtacttcagaggtatgaaacttgcaggctttt 60

QY 2859 aacaagaccatctacaaaagaccagattgtgataaacaataataacactaatagaa 2918
Db 61 aacaagaccatctacaaaagaccagattgtgataaacaataataacactaatagaa 120

QY 2919 gaaatgttccaggctcctcatctatatttgggtggcgttattgacctggagtggaat 2978
Db 121 gaaatgttccaggctcctcatctatatttgggtggcgttattgacctggagtggaat 180

QY 2979 gtgaccaaagagaggtcacaaatgagaaatcattcacttgccttgcttgacccatg 3038
Db 181 gtgaccaaagagaggtcacaaatgagaaatcattcacttgccttgcttgacccatg 240

QY 3039 ccacacagtgccattgcacaaaatttactgagaataataataaactgactgcttagag 3098
Db 241 ccacacagtgccattgcacaaaatttactgagaataataataaactgactgcttagag 300

QY 3099 aatgtcataaaacaaagtggccacattttaagaacaccaggtgtatcaggccatggagttat 3158
Db 301 aatgtcataaaacaaagtggccacattttaagaacaccaggtgtatcaggccatggagttat 360

QY 3159 gaactaaaagatgaatcactgaaagactccaatgatctacttttattcattactccaaaacc 3218
Db 361 gaactaaaagatgaatcactgaaagactccaatgatctacttttattcattactccaaaacc 420
```


Db 668 aagatacagacacccacacccggtgttgctccatctgcctctgtttgcaagcctgg 727
Qy 3331 ttaacctctcaactgtgatatactgatgtacattctcagagacogtatttgagcgga 3390
Db 728 ttaacatttgcagtcagatgtcatgttgcatactgggaacaattctgcaatggcgtg 787
Qy 3391 tagacacagattcaactgttgacaggaaggatgctcccaaatgcttttcaattctgg 3450
Db 788 tgaacataatggatagctgttgcagagtcctatgcaggaagggtttacatttaattg 847
Qy 3451 cattgggtttactagaagagaacacagcttcaaaaagctcctgaag---aagaagtaa 3507
Db 848 gcatggcactacaagaagaacacacatttagagagtgtaacggagagcatgtagtaa 907
Qy 3508 catttgactttatcataagcttcaagattgggaagtctcagccatgaat-----a 3558
Db 908 catttaccttcaacagatatacaaaacctgttgagcgccaaaaatttctctagca 967
Qy 3559 tacaaatgttttggaaaaactcaaggaattccccggttagaaggccagaagacatga 3618
Db 968 tactagctatgctggaaacactcaaaaatgctcctacctcctagaaagtcacaaaagacatga 1027
Qy 3619 taacgtgatacttcagattttgacacagtgaaagcgaatgaagagaaaaattctgtttaa 3678
Db 1028 ttcgggtggaattgagagactttcaatgtcttataaaagatgagggaga-----gttcaac 1081
Qy 3679 ttgtagcaaccacatcaggatcggaattcttaagaatgatgagattactcatgataaag 3738
Db 1082 ctaccagtcctgtggcagacagagaacacataatggaagagatttcaagggaacaaag 1141
Qy 3739 aaaaaacagacaaaagaaagctgaagctgtagctgtagctacatgcgcagaagatcatgg 3798
Db 1142 acaaaagtgaagagaagaaagacagatttgcagactgctgcagagaaagatcatgg 1201
Qy 3799 ctcaagatgctccttacagaaaaaacttcaaaactataaactcaactcatgtatgaacata 3858
Db 1202 ctcaagatgctgaaatgcagcgccattttattatgatgaaaaaagaactcttcagcaga 1261
Qy 3859 catcagaaaatgctgggaagagattccattatggaggaagagacacccacagcagtcga 3918
Db 1262 cattgaagctggatgcctcaacctgtgttcttgata-----tagcctgtggcctt 1315
Qy 3919 gtgactactctagaattgtttgggttcttaaacgggttccatctgtttactgaaaagagg 3978
Db 1316 cagatatgaacttacagcaactgggcccgcacaaaactcaggttccctgaaacaaagacaaat 1375
Qy 3979 tgcgtgagctgcatcctttgccagaagaacagaggtgaaaaatagaaaaataatgccatgg 4038
Db 1376 tegtacatgtaattgtgtcagaagagcaagaagttaagtggaaagcaggccaatgg 1435
Qy 4039 tattatcggcgtgtgtccagaagaacttactgccttaaccagcagacaggggaaccccatag 4098
Db 1436 tcttggcagcattgttccagagatacactgtattatacaaaaacagaagtaaaattttatc 1495
Qy 4099 aactctcagagagaagccttagacccttttcatggatccagacttggcatatggaactt 4158
Db 1496 aagatccagaaaaatat---gatccattattcatgcaccctgatctgtctgttggaaacac 1552
Qy 4159 atacaggaagctgtgtcatgttaatgcacgcagtgctgtggcagaagatattttgaaactg 4218
Db 1553 acactagtagctgtggcacattatgcataatgcccattgttggcaagatttttgattccg 1612
Qy 4219 tacagctgagctctcagcgcgcatctcatgttgacctttt-----tgacttgg 4266
Db 1613 ttcaagctaaagacagcggaagcaacagagattacgcttacatacagcagctatgtatgag 1672
Qy 4267 aaagtggagaatactttgcccctttttgcaaatctctgtgcaatactgtgatccccatta 4326
Db 1673 aaacaggaagaattcctttgcccctttgtgaatgcttgagtaataactgttatctctcgc 1732
Qy 4327 ttcctttgcaacctcaaaagataaaacagtgagaatgcagatgctctgtcctcaactttga 4386

Db 1733 tgcttt-----ctcaagaaaaatatattttaaacacaggtttaaaatttttctcagaccaacaa 1786
Qy 4387 ccttggcagcgttggatcacagactgtttctggccagaatatcatcaggtttaataataagacatg 4446
Db 1787 atctgactcagtggaatagaacaatatctcagcaataaaaagcattacaggtttcttagga 1846
Qy 4447 ctaaaaggagaaaaacccaattcctattttctttaaatacaaggaaatgggagattctactttgg 4506
Db 1847 aagaagaagtagtctcctaataatgcctctacaagaattcagaaaaatgttgatgaattac 1906
Qy 4507 agttccattccatcctcgtagtttggcgttgagcttcgatttaataatctcaaaatagcatca 4566
Db 1907 agtccctgaagggttcagggcctgatttctgtcctaagaaccttattctctgagagcatcaa 1966
Qy 4567 aggaataggtattctcttgcacacaacaaatttatagaatttgattgaaagtgcacactg 4626
Db 1967 aagaatgctaacgacatttggaaactgctacotacaaggtgggactaaaagttcatccca 2026
Qy 4627 atgaaggagatcctcagatcccatgctgacotcggagacactgcgtcttcaactaccag 4686
Db 2027 atgaaggagatcctcgtgtcccataatgtgttgggttagctgcgtcagcaccatccaaa 2086
Qy 4687 caattgaaaaatctattgggagatgaaggaacacctctgttctgagcacttcaaaaatagcc 4746
Db 2087 gcatagaagaatatttgatgatgaagataaacacattgttctgcttacccttgcagac 2146
Qy 4747 agcataataggttgaaagcatttaatacagtttgcagttgcacagagagattacctgtcctc 4806
Db 2147 tggatgactgtctttagtcatctgacgagatttgcgcagcacactggacagtggtcatcag 2206
Qy 4807 aggtcctgatacacagaacacatcgttctgtcttctatacagttgttcttcttaacataaaat 4866
Db 2207 ttctagtggtgcaagacattttttaaacttttgcactcactcgttgccttaatgacagcc 2266
Qy 4867 cagaagatacacacgccttctgtctatagatcgttttcaatgttttgggtgggtgctgtgt 4926
Db 2267 atggaagacttccatgcagcattagatcattgacatgtttcatttatgggtgggttggcgc 2326
Qy 4927 tagcattcccatccttctgtattgggagacccgttgcagccttctcagtttagtt 4986
Db 2327 ttgcatttctcgtgtgcaagt-----tcaggatttttcagggaatcagcc 2371
Qy 4987 ctctctataaccacaccttttctcttccatttgatcacaccatggcacacactgcttcagatcc 5046
Db 2372 ttggcactggagaccttccatttccatctgtgttactatggcacatcatcacagatct 2431
Qy 5047 tacttacagtagaacac---aggcctacccttgcctcaggttcaagaagacagtgaaagg 5103
Db 2432 tacttacctcatgtacagaagaaatggcatggatcaagaaaatcccccttctggaagaag 2491
Qy 5104 ctcatccgcgcatcttcttcttctgcagaaatttctcaatatacaagtggtcctcattgggt 5163
Db 2492 aatcagcagttctcgtttgtataaaaacacttccacagtatcacgggaagtgccttgaaag 2551
Qy 5164 gtgatattcctgggtgtatttctgtgtcactgaagaatgggcatcaccccttatcttc 5223
Db 2552 aaatcaccatccggctggcactctgtggaggagtgacagcgtggaatcatccttctctga 2611
Qy 5224 gctgtgcgcatgttttctcactatttacttgggttaactccgcctcaggaactgcata 5283
Db 2612 agtgtctgcttatttttcttacttaaatggagttctctcccccacccagatccaa 2671
Qy 5284 ccaattctgcagaagagagtagacagtgcaactctgtagctatctatcttaccctacaaatt 5343
Db 2672 t---tcttggaaacagccatttgaacatttatgtatgtatcttctccctacaaacaaac 2728
Qy 5344 tgttctcgtcttctccaggaaatattgggatactgtaaggcccttgcctccagaggtgtgtg 5403
Db 2729 tcaattgctcttttcaagaaaaatagtgagataatgaattcactgattgaaagtgtgtgccc 2788
Qy 5404 cagatcctcgtccttactaaactgtttgaaagcaaaaaaacccgttggcaggtacccctagaa 5463
Db 2789 gtaacagtgaaagttaaaagatactcagaaaggtgaaagagatgctataagaatatccaaag 2848

QY 3391 tagacacagattctaaacttggaccggaagggtgctcccaaatggccttttcattattcgg 3450
D 3451 cattgggtttactagaagaagaacacagcttcaaaaagctcctgaag---aagaagtaa 3507
D 903 gcatggcactacaaagaagaacacacatttagaagatgcacggaagcatgtagtaa 962
QY 3508 cattgacttttatcataagcttcaagattggaagttggaagttcagccatgaat-----a 3558
D 963 catttaccttccacagaagatcatcaaaacctggtgaagcgcacaaaatttccctagca 1022
QY 3559 tacaaatgcttttggaaaaactcaaaaggaattcccccagtttagaagccagaaagacatga 3618
D 1023 tactagctatgctggaacactacaaaagtctccctaccttagaagtcaccaaagacatga 1082
QY 3619 taacgtggatacttcagatgtttgacacagtggaagcattaaagaaaaatcttgtttaa 3678
D 1083 ttcggtggatattgaagacttttaactgctttaaaaagatgagggaga-----gttcac 1136
QY 3679 ttgtagcaaccacatcaggatcggaatctattaaagatgatgagattactcatgataaag 3738
D 1137 ctaccagtcctcgtggcagagacagaaggaacccataatggaagagagttcaaggggacaaag 1196
QY 3739 aaaaagcagaacgaaaaaagaaagctgaagctgttaggtctacatgccagagaagatcatgg 3798
D 1197 acaaaactgagagagaagaagacagagattgccagactgcgcagagaaaagatcatgg 1256
QY 3799 ctoagatgctccttacagaaaaaactcattgaaactcataaactcatgtatgacaata 3858
D 1257 ctcaagatgctgaattgcagcggcattttattgtatgataaaacaaagaactctttcagcaga 1316
QY 3859 catcagaagatgctggaaagaagattccattatgaggaagaagacaccccagcagtcga 3918
D 1317 cattagaagctgagtcctcaaacctctgctgttcttgatca-----tagccctgtgctt 1370
QY 3919 gtgactactctagaattgttttggctccataaacgggggtccatctgttactgtaaaaggagg 3978
D 1371 cagatatgaacttacagacactgggcccgcacaaactcaggttcctgtaacaaagacaaat 1430
QY 3979 tgctgacgtgcatcctttgccagaagaacagagagtgaaaaatagaaaaatagccatgg 4038
D 1431 tcgtttacatgtatattgttcaagaggagcagaagttaaagtggaaagcagggcgaatgg 1490
QY 4039 tattatcgctgtgtccagaataactactgccttaaccacagacagaggaagaaacccatag 4098
D 1491 tcttggcagcattgttcagagatacaactgtattatcaaaaaacagaagtaaaatttattc 1550
QY 4099 aactctcaggagaagccctagaccacttttcatggatccagacttgcataatggaaactt 4158
D 1551 aagatccagaaaaaat---gatccattattcatgcacccctgatctgtcttgggaaacac 1607
QY 4159 atacagggaagctgtggtgtcatgtaatgacgcagtgctggtggcagaagttatttgaagctg 4218
D 1608 acactagtagctgtgggcaactaattgcatgcccattgttggcaaaagttatttatttgcg 1667
QY 4219 tacagctgagctctcagcagcgcattcatgttgaacctttt-----tgacctggg 4266
D 1668 ttcaagctaaagaacagcgaagggcaacagagatatcgctttacatacagactgatgtatg 1727
QY 4267 aaagtggagaatatcttggccctctttgcaaatctcttgcaatactgtaactcccatcta 4326
D 1728 aaaaagagaaatcttcttggccctcttggaaatgcttggaaatgtaactgttattctctgc 1787
QY 4327 ttcctttgcaacctcaaaagataaaacagtgagaatgcagatgctctgtgcacatttga 4386
D 1788 tgcctt-----ctccaagaaatatttttaacaaacaggttaaatttttcagaccaccaa 1841
QY 4387 ccttggcagctgtggatagacagctgttctggcagaaatcatcaggtttataataaagacatg 4446
D 1842 atctgactcagtggaattagacaaatctctcagcaaaataaaagcattacagtttcttagga 1901
QY 4447 ctaaaaggaaaaaccccaattctctattttctttaaatacaagggaatgggagatttacttgg 4506

D 1902 aagaagaagtagtactcttaataatgctctcaaaaattcagaaaaatgtggatgaattac 1961
QY 4507 agttccattccatcctcctgagtttttggcgttggagttcttcgatttaataatctcaaaatagcatca 4566
D 1962 agtccctgaagggtttcaggcctgatttgcctcaagatcccttattcttgagagcataa 2021
QY 4567 aggaataatgtttattctcttggccacaaatattatagaatttgatgaaagtgcacactg 4626
D 2022 aagaataatgctaaacgacatttggaaactgctacactacaaggtgggactaaaagtctatccca 2081
QY 4627 atgaaaggaatcctcaggtcccatcgtcactgcggagcactgcgctttcactatccagg 4686
D 2082 atgaaaggaatcctcgtgttcccataaattgttgggttagctgcgctacacattccaaa 2141
QY 4687 caattgaaataatattgggagatgaaggaaacacctctgtttggagcacttcaaaaataggc 4746
D 2142 gcatagaagaatatttgatgatgaagataaaacctattgttggctcttacccttgagac 2201
QY 4747 agcataatggtctgaaagacattatgcagtttgcagttgcacagagagattaccctgctc 4806
D 2202 tggatgactgtcttaggtcattgacgagatttgcgcagcacaactggacagtggcatcag 2261
QY 4807 aggtcctgatatcagaaacatctcgttctctctatcagttgttcttcttaacataaaaat 4866
D 2262 tttcagtggtgcaaggacattttttaaacttttgcataactgttgcataatgcagacc 2321
QY 4867 cagaagatacacaccatgcctctctgtctatagatctgtttcatgttttgggtgggtgctgtgt 4926
D 2322 atgaggaacttccatgcataattagatatattgacatgtttcattatttgggtggtgctg 2381
QY 4927 tagcattcccatcctgtgtattgggtgagatgacctgttgcagctcttctcagtttagtt 4986
D 2382 ttgcatcttctcgttgcagt-----tcaggatttttcagggatcagcc 2426
QY 4987 ctctcctaaacacacctttatctcttccatttgcacatgcacacacatgcttccagatc 5046
D 2427 ttggcacttggagacctcacaatttccatctggttactatggcacacatcacagatct 2486
QY 5047 tacttactagtagacac---aggcctacccttgcctcaggttccaagaagacagtgaaaggg 5103
D 2487 tacttactcctatgcacagaagaagaatgcattgatacaagaatacccccttggaaagag 2546
QY 5104 ctaattccgcactcttcttcttgcagaaatttctcaatatatacaagtggtccattgggt 5163
D 2547 aatcagcagttcttgccttggataaaaacacttccacagtatatacgggaagtgccttgaag 2606
QY 5164 gtgatatctcctcgtgtatttgggttgcactgaagaatggcactaccccccttattcttc 5223
D 2607 aatcaccatccggtggtgcatctgtggagagtgctcagagctggaatcatgcttcttctga 2666
QY 5224 gctgtgctgcatgtttttccactatttacttgggttaactccgctgaggaactgcata 5283
D 2667 agtgtctgtcttatttttcttacttacttaaatggagtcttccccccacacacacag 2726
QY 5284 ccaattcgcagaaggagagtacagtgcaactgttagctgtatctatttactacataaaatt 5343
D 2727 t---tcttggaaacagccatttgaacatttatgtagctatttcttccctaccacaaaccc 2783
QY 5344 tgttctcgtcttccaggaataattgggatactgttaaggcccttgcctcagaggtgtgtgtg 5403
D 2784 tctatttgccttttcaagaaaaatagtagataatgaatttcaactgattgaaagtgtgtgc 2843
QY 5404 cagatcctccttactaaactgtttgaagcaaaaaaacacccgtggtcaggtaccctagaa 5463
D 2844 gtaacagtgaaagttaaagaatctagaaggtgaagagatgctataagatcccaagag 2903
QY 5464 aagaataatgtttgatagagcttctcctgactatagctatagctgcctcctggaatcaagcttctc 5523
D 2904 aatctaacaataataaaaccttccagaggattacagcagcctcatttaatacaagcatcca 2963
QY 5524 atttccaggtgcacacggctgcagatgatgagcgaagacatcctgtcctcgtccttttct 5583

Dbb 2964 atttctgtgccccaaatacaggtgtgtataagacagagcccccaactctgtgccttgtgt 3023
Qy 5584 gtggggtctactatgttctcagacaatttctgcccagaaattgtgaacggggaagagg 5643
Dbb 3024 gcgaatctgtgtgtccagatttactgtctccagactgaactggaagggaagatg 3083
Qy 5644 ttgagctgttttttccagcacttcaactgtgagccggaggtctgtatttctctaaaaa 5703
Dbb 3084 taggagctgcacagctcacacctactctgtgctctggtgagtggtcctctcctgagag 3143
Qy 5704 tcagagaatccagtggtctgtgtgaagttaagtaaacccagagaggtgtccctatccagctc 5763
Dbb 3144 taccggaatgcaggtgtattttagctggcaaaccaaggctgtttttattctctc 3203
Qy 5764 ctacttgatgaatatggagaacagacctggcctggaagaggggcgaaccccttcatt 5823
Dbb 3204 ctactctgatgactatgggagaccagcaggactcagagcggggaaatcctttacatt 3263
Qy 5824 tatctgtgagcggtatcggaagctccatttggctctggcaacaacactgcattatagaag 5883
Dbb 3264 tatgaaagcaggtattcaagaagattcagaagctctggcaccacaacacagtgctcacagag 3323
Qy 5884 agattgttagagcgaagagactaatcagatgtttatttggattcaactggca 5935
Dbb 3324 aaattgacatgcacaggaagccaatcagacactggttggtgactgactggca 3375
RESULT 8
AAA02411
ID AAA02411 standard; cDNA: 733 BP.
XX
AC AAA02411;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2402.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 21-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEO INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR WPI; 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX
PS Claim 1; Page 958; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present

CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotide sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 733 BP; 218 A; 145 C; 152 G; 210 T; 8 other;

Query Match 10.4%; Score 656.8; DB 21; Length 733;
Best Local Similarity 97.5%; Pred. No. 7e-169;
Matches 696; Conservative 0; Mismatches 13; Indels 5; Gaps 3;

QY 1405 catatgaccacgtcatatcacagcctacaaagagctcttgactgtgactgcagagggccc 1464
Dbb 20 cagcagaaccagtcatacacagcctacaaagagctcttgactgtgactgcagagggccc 79

QY 1465 agttgcataccactgccattgacaaagaggtgctgcgggtgttaaaagcgggagctcatg 1524
Dbb 80 agtgcnttcactgcattgacaaagaggtgctgcgggtgttaaaagcgggagctcatg 139

QY 1525 ctgcttgccaggaagcaaaagagataaaagagtcattcagaaaaatgtctcacaacatc 1584
Dbb 140 ctgcttgccaggaagcaaaagagattttaaagagtcattcagaaaaatgtctcacaacatc 199

QY 1585 cacttcgatagaagtattacactcagagattatggtctcatcagaaattgctttgcgtc 1644
Dbb 200 cacttcgatagaagtattacactcagagattatggtctcatcagaaattgctttgcgtc 259

QY 1645 ttggttcctggatgaacaaaattatgacttatcagtgacttagcagatctttgccc 1704
Dbb 260 ttggg-tcctggatgaacaaaattatgacttatcagtgacttagcagatctttgccc 318

QY 1705 aagcatgccttagagaagaacacctgactcggagaaatccctctcacaagcaggttaatgc 1764
Dbb 319 aagcatgccttagagaagaacacctgactcggagaaatccctctcacaagcaggttaatgc 378

QY 1765 ttgggatgcaaaagctttataaaaggtgccgttaagatcccttcattgaattgatctcagca 1824
Dbb 379 ttgggatgcaaaagctttataaaaggtgccgttaagatcccttcattgaattgatctcagca 438

QY 1825 gttttttatgagatggaatacaaaaaactcttgcctatggaatttgcgaagtattata 1884
Dbb 439 gttttttatgagatggaatacaaaaaactcttgcctatggaatttgcgaagtattata 498

QY 1885 aacaactgcagaagaataatatacagtgatcatcagacagagtatctctataactcac 1944
Dbb 499 aacaactgcagaagaataatatacagtgatcatcagacagagtatctctataactcac 558

QY 1945 ttctgattcagatgtttactgttcctactctgctgcagactcttatttgaagacagagt 2004
Dbb 559 ttctgattcagatgtttactgttcctactctgctgcagactcttatttgaagacagagt 618

QY 2005 ttatctctgattactgaaactctgtagaagtatttaccctgactgactggacagg--aa 2062
Dbb 619 ttatctctgattactgaaactctgtagaagtatttaccctgactgactggacagg--aa 678

QY 2063 caataaatcaacttccagggttatagcc--aggacaaattgggaagagtatat 2114
Dbb 679 caataaatcaacttccagggttatagccaggcagggaacaaattgggaagagtatat 732

RESULT 9

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:39:35 ; Search time 4409.56 Seconds
(without alignments)
23599.681 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattcgccacgagg.....aatttgattgtgtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4435.4	70.3	6395	6	AR030784	AR030784 Sequence
2	4435.4	70.3	6395	6	AR121463	AR121463 Sequence
3	4435.4	70.3	6395	10	AF061555	AF061555 Mus muscu
4	2546.8	40.4	2550	9	AK027803	AK027803 Homo sapi
5	996.2	15.8	1001	6	AR030785	AR030785 Sequence
6	996.2	15.8	1001	6	AR121464	AR121464 Sequence
7	994.2	15.8	999	9	AF061556	AF061556 Homo sapi
8	981	15.6	6158	9	AB002347	AB002347 Human mRN
9	764.2	12.1	818	9	HSX521	Z47040 Human part1
10	651.6	10.3	2958	9	AC026998	AC026998 Homo sapi
11	503.8	8.0	166518	9	AC090514	AC090514 Homo sapi
12	487.8	7.7	2512	9	AK026948	AK026948 Homo sapi
13	354	5.6	190727	9	AC068724	AC068724 Homo sapi
14	279.4	4.4	61896	2	AC090992	AC090992 Homo sapi
15	215.2	3.4	910	10	MM1UBR2	AF067372 Mus muscu
16	212.2	3.4	107304	2	AC016274	AC016274 Homo sapi
17	212.2	3.4	164468	2	AC021899	AC021899 Homo sapi
18	180.8	2.9	164468	2	AC021899	AC021899 Homo sapi
19	160.6	2.5	3980	9	HS1UBR4	AF067383 Homo sapi
20	150.2	2.4	107304	2	AC016274	AC016274 Homo sapi
21	132	2.1	28834	2	AC017982	AC017982 Drosophil
22	132	2.1	152545	3	AC010922	AC010922 Drosophil
23	132	2.1	179016	3	AC018489	AC018489 Drosophil
24	130.4	2.1	900	9	HS1UBR3	AF067382 Homo sapi
25	123.4	2.0	907	10	MM1UBR1	AF067371 Mus muscu
26	115.2	1.8	61896	2	AC090992	AC090992 Homo sapi
27	111.8	1.8	2616	10	MM1UBR6	AF067376 Mus muscu
28	106.4	1.7	830	9	HS1UBR5	AF067384 Homo sapi
29	104	1.6	143291	9	AL137792	AL137792 Human DNA
30	104	1.6	143697	2	AL358572	AL358572 Homo sapi
31	104	1.6	295265	2	AL353354	AL353354 Homo sapi
32	102.6	1.6	716	10	MM1UBR5	AF067375 Mus muscu
33	101.8	1.6	1000	9	HS1UBR6	AF067385 Homo sapi
34	100.8	1.6	1370	9	HUMORF02	D14658 Human mRNA
35	100.4	1.6	1100	10	MM1UBR4	AF067374 Mus muscu
36	97	1.5	179588	2	AC009852	AC009852 Homo sapi
37	97	1.5	182660	2	AC012416	AC012416 Homo sapi
38	97	1.5	183982	2	AC009825	AC009825 Homo sapi
39	91.4	1.4	1000	9	HS1UBR7	AF067386 Homo sapi
40	91.2	1.4	727	10	MM1UBR7	AF067377 Mus muscu
41	89	1.4	789	4	CFU12687	U12687 Canis famil
42	88	1.4	168639	2	AP001992	AP001992 Homo sapi
43	88	1.4	213613	2	AC013405	AC013405 Homo sapi
44	86.4	1.4	400	9	HS1UBR1	AF067380 Homo sapi
45	86	1.4	195064	2	AL138921	AL138921 Homo sapi

ALIGNMENTS

RESULT	1				
AR030784					
LOCUS	AR030784	6395 bp	DNA	PAT	29-SEP-1999
DEFINITION	Sequence 1 from patent US 5861312.				
ACCESSION	AR030784				
VERSION	AR030784.1	GI:5943998			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 6395)				
AUTHORS	Varshavsky, A. and Kwon, Y. Tae.				
TITLE	Nucleic acid encoding mammalian UBRI				
JOURNAL	Patent: US 5861312-A. 1 19-JAN-1999;				
FEATURES	Location/Qualifiers				
source	1..6395				
BASE COUNT	1802 a 1376 c 1519 g 1698 t				
ORIGIN	/organism="unknown"				

Qy 4897 atctgttctatgttttgggtggtgctgtagcattcccatctcttcttatttggatgacc 4956
Dy 4328 ATCTCTTCCATGTTCTGGTGGGCGAGCTTAGCGTTCCCATCTCTTATTGGGATGACA 4387
Qy 4957 ctgttgaatcgcagccttctcagtttagttcttccataaccacctttatcttccatt 5016
Dy 4388 CCGTGGATCTGACGCGTCCGCTTACCTTATGTTCTTATATACCCCTCTATCTCTCCATC 4447
Qy 5017 tgatcacatggcacacatgcttcacatactactactacagtagaca-----cag 5064
Dy 4448 TGATCACATGGCGACATGCTTCAGATACTCTTACACAGATACAGATGCTGCTCCAG 4507
Qy 5065 gctaccccttgcaggttccaagaagacaggtgaagaggctcattccgcattcttcttct 5124
Dy 4508 GCGCGCGCTTGTGAGGGTGAAGAGATAGTGAGGAGGCTCGTGTGCATCTGCTTCT 4567
Qy 5125 ttgcaaaaatttccaatatacaagtggctccattgggtggtgatattccctggctgatt 5184
Dy 4568 TTGTGGAGTGTGCGACACACAGCGGCTTCACTGGGTGGGTGCTCCGCGCTGGTACC 4627
Qy 5185 tgtgggtctcaactgaagaaatggcatcaccccttattctcgtgctgcatgtttttcc 5244
Dy 4628 TGTGGCTCTCCCTGAGAACCGGATCACCCCTTACCTCCGCTGCTGCTGCTTTTCC 4687
Qy 5245 actatttacttgggttaactccgctgaggaactgcataccaaattctgcgaagagagt 5304
Dy 4688 ACTATTTACTTGGAGTAGTCCGCGCTCAAGAACTGTTTGCAATTTCTGCTGAAGGAGAT 4747
Qy 5305 acagtgcactctgtagctatctattcttaccatacaaaatttgttctcctctccaggaat 5364
Dy 4748 TCAGTGCACCTGTGTAGTATCTATCTTTACCCACAAATTTGTCTCTCTTTTCCAGGAAT 4807
Qy 5365 attgggaactgtaagggcccttgcctcagaggtggtgagatccctgccttactaaact 5424
Dy 4808 ATTGGGATACATAAGGCCCTTACTACAGAGGTGGTGTGGAGATCCTTACTCAAGT 4867
Qy 5425 gtttgaagcaaaaaacacccgttggtcaggtacccctagaaaaaagaaatagtttgatagc 5484
Dy 4868 CTTTGAAGCAGAAAGTGTGTGTGTCAGGTACCTAGAAAAGAAATAGTTTGATAGAGC 4927
Qy 5485 ttctgtagactatagctgctcctcctgaatcaagcttctcattcaggtgcccaggtctg 5544
Dy 4928 TTCTTGAGGACTACAGCTGCTTCTAAATCAGGCTTCTCACTTAGTGTGTCACGGTCTG 4987
Qy 5545 cagatgatgcaaaagcattcctgtcctcctccttctcctgtggggtactatgttctc 5604
Dy 4988 CAGATGATGACGAAACATCTCTGCTCTCTCTTTTCTGTGGGCCATCTGTGTCTC 5047
Qy 5605 agaactttgtccaggaaatttgtgaacggggaagaggttgagcttgcatttttcacg 5664
Dy 5048 AGAACATCTCTTGCCAAAGAAATAGTGAATGGGAAGAGGTGGAGCGTGGCTTTTTCATG 5107
Qy 5665 caattcactgtgagcggagctgctgatttctcctaaaaacagagaatgcccaggtgctc 5724
Dy 5108 CGTTTCTTGTGTGTGAGTGTGATTTTCTTAAATCCGAGAAATGACAGGTTGTCTC 5167
Qy 5725 tgggtgaagtgaagcagaggtgctgctcctcctcagctccttacttgatgaatatgag 5784
Dy 5168 TGGTGAAGAAAGAACGACAGGCTGTGCCATACCCAGCCCTTACTTGGATGAATATGAG 5227
Qy 5785 aaacagaccctggcctgaagagggggcaaccccttctatttctcgtgagcgggtatcgga 5844
Dy 5228 AACAGACCCAGGCTAAACAGAGAGAAACCCATTCATTTATCTCGGAGCGGTATCGGA 5287
Qy 5845 agctccattgttggcgaacaacactgcatttatagagagattgctagagcgaagaga 5904
Dy 5288 AGCTGATTTTGGTGTGGCAACAGCATGCTATATAGAAAGATATGCTCGGAGCCAGAGA 5347
Qy 5905 ctaatcagatgttatttgattcaactggcagttactgtgaactccaactcctcctcaag 5964
Dy 5348 CTAATCAGATGCTATTGGGATTTAACTGGCAGTTACTCTGAGCTTCAAGTTGTGCTCAAG 5407

Qy 5965 acaatcacaaatgacagcagtagtaaaaggctgattcaaaaattatggaaaaacttctgagg 6024
Dy 5408 ACAATCATGAGTGACATCAATAATAAAGACTGATCTAAATTTCTAGAGAATTTCTGAGG 5467
Qy 6025 gctggaaagtattgagggctcttttgcctatgctccaggttccactatcatcaataaaat 6084
Dy 5468 ACGGGGAAGTATTGGAGGGTCTTTTGATCCATGTCAGATTCACACACATTAATAAAT 5527
Qy 6085 atttctaatggagtattgcttctcaattagcaaaacatatgcttccacagaaaaa-aggac 6143
Dy 5528 ATTCTCTAATGCAATATTGCTTTCAATTATCAACATAAGCTTCAAGGGGAAAAACAAGAC 5587
Qy 6144 atagatacaatctgt-----ttt 6160
Dy 5588 ATAGATTAAATGTTTATGTTCTTAGAACACATAAAGAAATGCTTTGTCATCCAAAGTGCTAT 5647
Qy 6161 atgtctagtatttccagaaatttattcccttcaataatttgcctcatttcttatttattt 6220
Dy 5648 TTCTGCTAATATTCCAGAAAACCTCTTTCCCTTCATACTGTCCTAGTTCATTTTCATAT 5707
Qy 6221 catccacttggtagaagtcacgtcacaacagttgtagacatttttatgtttggttaac 6280
Dy 5708 CACCACCTGGTTAATGAGTCAATTAAGCATTTGTGGACATTTCTCCATCTGGCTAAC 5767
Qy 6281 tcttctgcaattttgtatttgggtttt 6307
Dy 5768 ATCTGACCTTTGTTATTGTTGTTT 5794

RESULT 2
AR121463 LOCUS AR121463 6395 bp DNA PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6159732.
ACCESSION AR121463
VERSION AR121463.1 GI:14105039
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6395)
AUTHORS Varshavsky, A. and Kwon, Y. Tae.
TITLE Nucleic acid encoding mammalian Ubri
JOURNAL Patent: US 6159732-A 1 12-DEC-2000;
FEATURES Location/Qualifiers
source 1. 6395
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

Query Match 70.3%; Score 4435.4; DB 6; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

Qy 590 ttcagggggccgtcgttaaagtgtcctcctgtctc-tccgaccggccacaggtttccgt 648
Dy 8 TTCAGGGGCGCTCGTAAAGTGTCTCCTGTCTGCGTGTGGCGCCACAGGTTTCCGCT 67
Qy 649 tgcctctggccgggggtcggaactgcaggcgtcagtttccctcaagatgaggcagag 708
Dy 68 AGCTGGCGCGGGGTCTGGGAACCTGCGGGGCTGTCTTTCCCTTAAGATGGCGGACGAAG 127
Qy 709 aggtcggaggctactgagaggtatgaaatcagcgcgaggttaccaccagaccctcagcgtc 768
Dy 128 AGATGGACGGCCCGCAGAGGATGGACGTACGCCCGGAGCCTCCCTGGCCCGCAGCGGC 187
Qy 769 tggcatcttgggtggatcagcaagtgtgatttttatactgcttcttgcattcttggcac 828
Dy 188 CGCATCTGTTGGGATCAGCAAGTTGATTCTATATCTTCTTACATCATATTGGCAC 247
Qy 829 aattggtccagaaatttactttgctgtaaatggacccagacttgaaaaagcggagagaaa 888
Dy 248 AATTAGTCCAGAAATTTATTTTGTGATGGACCCAGATTGGAAAAGCAAGAGAGA 307

Qy 889 gtgtacaaatgtcaaatattcaactccactgaaatgggtacttatttgggaagatccagata 948
Db 308 GTGTACAGATGTCAATACTCACTCTTGGAGTGGTACTTATTTGGAGAGATCCGGATA 367
Qy 949 tttgttagagaaatgaagcagatgagcatttcagcttgggtggaggttttcaaaa 1008
Db 368 TTTGCTTAGAANAATTAACACACAGTGGAGGTTCAGTGTGTGGGAAGGTTTCAAAA 427
Qy 1009 gtggagagacaactattcttgcagggatgtgcaattgatccaacatgtgtactctgta 1068
Db 428 GTGGAGAAACAACATATTCCTGTAGGGATTTGCAATGTATCAACCTGTGTGCTCTGTA 487
Qy 1069 tggactgtccaggagcagtggttcaataaaatcogttacaagatgcatacttctactg 1128
Db 488 TGGACTGCTTCCAAAGTAGTGTTCATAAAAACCATCTTACAAGATGCATACTTCTACTG 547
Qy 1129 gaggaggttctgtgactgtggagacacagaggcatggaaaaactggccctttttgtgtaa 1188
Db 548 GAGGGGCTTCTGTGACTGTGGAGACACAGAAGCGTGGAAAACCTGGCCCTTTTGTGTGG 607
Qy 1189 atcatgaacctgtgaagacaggttactataaagagaattcacgctgtccgttgaatgaag 1248
Db 608 ATCAGAGCCTGGAAGACAGGTACTACAAAGAGAGCTTTACATTGCCCATTTGAATGAAG 667
Qy 1249 aggttaattgtccaggcaggaataatttctctcagtgataaaatatgtcgtgataaatga 1308
Db 668 AGGTGATTGTCTAAGCCAGCAGAGAATATTCCTTCGGTGTATAAATACATTGTAGAAATGA 727
Qy 1309 ctatatgggaagggaagaaactgcctcctgaactccagataaaggagaaatgaaa 1368
Db 728 CTATATGGGAAGAAAGAAAGAAATGTGCTCTGAACCTGCAGATTAAGGGAGAAAAATGAAC 787
Qy 1369 gatactattgtcccttttcaatgaatgaacacacatccatgaccacgtcatatacagcc 1428
Db 788 GATACTATTGTGCTTTTCAACGATGAGCACCATTTCGTATGATCATGTATCTACAGTC 847
Qy 1429 tacaagagctcttgactgtgagctgcagaggccaggttgcataccactgccaattgaca 1488
Db 848 TGCAGAGCTCTAGATTGGAGCTTGCAGAGGCACAGCTGCACACACTGCCATCACA 907
Qy 1489 aagagggtcgtgggtgttaaagcgggagcttatgctgcttggccaggaagcaaggaag 1548
Db 908 AAGAGGTCGCCGGGTGTCAAAAGCAGGTGTGTATGCCACTTGGCCAGGAAGCAAGGAG 967
Qy 1549 atataaagatcattcagaaaatgtctcacaacatccattcatgtagaagtattacact 1608
Db 968 ATATAAAGATCACTCAGAAACGTCTCTCAGCACCCCTCCATGTGGAAGTGTGCAC 1027
Qy 1609 cagagattatggctcatcagaaaatttggcttgggtcttggcttcctggatgaaacaaatta 1668
Db 1028 CCGTGGTTATGGCTCACCAAAATTCGCTCTGCCCTTGGCTCTGGATGAACAAATTA 1087
Qy 1669 tgaacttaaatgacttttagcagatcttttggcaagcatgccccttagaagaacactg 1728
Db 1088 TGAGCTATTCAAGTGACTTTAGACAGATATTTTGGCCAGGCTTGCCCTCTAGAACAACTG 1147
Qy 1729 actcgagaatccctgtctcataagcaggttaatgcttgggatgcaagcctttataaag 1788
Db 1148 GCTCTGAAAATCCCTGCTTATAGCAGACTAATGCTTTTGGGATGCAAAACTTTTAAAG 1207
Qy 1789 gtgcccgaagatcccttcagaaatttgatcttcagcaggtttttttatgtgagatggaataca 1848
Db 1208 GTGCCGTAGATCCCTCATGATTTGATCTTTAGTAGTATTTTATGGAGATGGAATACA 1267
Qy 1849 aaaaactctttgtctatggaaatttgtgaagtattataaacaactgcagaaagaatatca 1908
Db 1268 AAAAACTCTTTGTATGGAATTTGTGAAGTATTATAAACAACCTGCAGAAAAGAGTACATCA 1327
Qy 1909 gtgatcatatgacagaagtatctctataaactgcacttccagttcagatgttactgttc 1968
Db 1328 GCGACGACCAACGAGAGAGATCTCCATAACCGCCCTGTCCGTCAGATGCTCACCCCTCC 1387

Qy 1969 ctactcttggctcgacatcttatttgaagacagaaatgtatctctgtcattactgaaactc 2028
Db 1388 CGACCTTGGCCCGGATCTTATTGAAGACAGAAATGTTATTCTGTCTACTACTGAAACGC 1447
Qy 2029 tgcctagaagtttactcagttacttggacaggaacaaataaattccaacttccaggggtata 2088
Db 1448 TGCTAGAGATTTTACTCTGAATACTTGGACAGGAACAATAAATTTCAACTTCCAGGGTTATA 1507
Qy 2089 gccaggaacaaattgggaagatgatgtcagtaatatgtgacctaaagtatatctctgatca 2148
Db 1508 GCCAGACAAACTGGGAAGAGTCTACGCCAGTATATATGTGACCTAAAGTATATCTGATTA 1567
Qy 2149 gcaaacccacaatatggaacagaaaattaaagaatgcagtttcccttgaaggttttcgactct 2208
Db 1568 GCAAGCCTGTATATATGGACAGAACGATTAAGAGCGCAGTTCCTGGAAAGGTTTCCGGTCTT 1627
Qy 2209 ttttgaagattcttaccctgtatgcagggaatggaaagaaatccgaagacaggttgggcaac 2268
Db 1628 TTTCTGAAGATCTTACCTGTATGCAGGGAATGGAAGAAATCAGAAGACAAGTTGGACAAC 1687
Qy 2269 acattgaagtgatcctgattgggaggtgccattgtctatcacagatgccaattgaagaata 2328
Db 1688 ACATTGAAGTGGACCCCTGACTGGGAGGCTGCCATCCTATACAGATGCCAACTAAAGAATA 1747
Qy 2329 ttttactcatgttccaaagatgggtgtgtgtgataaagaactcttacttgtggcttata 2388
Db 1748 TTTTCTCATGTTTCCAAGAGTGGTGTGTGTGATGAAGATCTCTTACTTGGTGGCTTATA 1807
Qy 2389 aagaatgtcaaaaagctgtgatgaggtgcagtcaccagtttcatatctagttagcaagacag 2448
Db 1808 AAGAAATGTACAAAGCTGTAAATGAGGTGCAGTACAAATTTTCATGTCCAGTACCAGACAG 1867
Qy 2449 tagtacaatcgtgtgacatagtttggaaacaaagtcctcacagatctatgagatcttg 2508
Db 1868 TAGTGCATTTGTGCGGTCTATAGTCTGGAAACCAAAATCTCACAAAGTCTCTGAGGACCTTG 1927
Qy 2509 taagcatatctgcactctcttagacccttqctgtgtcttcatgtacgttttaagcaggc 2568
Db 1928 TAAGCATACACTGCCACTCTCTAGAACACTTGCCTGTCTTCATGTACGTTTAAGCAGAC 1987
Qy 2569 tgggtgctgtttcaagactgcataaattgtgtcttcttttggaggacttccaagttagagttac 2628
Db 1988 TAGTGTCTATTTCAAAGACTGCATGAATTTGTGCTTTTGACAGCTTTCAAGTAGAGTCC 2047
Qy 2629 tagtgaatactcttactgtctgt 2688
Db 2048 TGGTGGAGTACCCGCTGCGCTGCCCTGGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 2107
Qy 2689 gaagaaatggactgtctcttattagccaggtgtttttattaccagaagtgttaagtgcagag 2748
Db 2108 GAAGAAACGGGCTCTCACTCATCAGCCAGGTTTCTTATATCAAGATGTTAATGCGAGG 2167
Qy 2749 aagaatgtatgaataagataatcatatgtcttcaagattgggtgcattcttaattgaatccca 2808
Db 2168 AGGAAATGATCAGATAAAGATATCATCATGCTTCAGATTTGGAGCATCTATAATGGATCCCA 2227
Qy 2809 ataagttctgttactgttacttcagagatgaacttgcagaggtttttaaacaaccca 2868
Db 2228 ACAAGTCTTGTGTACTGGTACTTCAGAGATATGAATCTTACTGTCTTTTAAACAAGACCA 2887
Qy 2869 tatctacaaaagaccaggaatttgaataaacaataataacactaatagaagaatgcttc 2928
Db 2288 TATCCACAAAAGACCAGGATTTGATTAACAGTATTAATACATTAATAGAGAANAATGCTTC 2347
Qy 2929 aggtcctcatctatattgtgggtgagcgttatgtacctggagtgggaaaatgtgacaaag 2988
Db 2348 AGGTCTCATCTATATTGTGGGAGAACGTTATGTACCTGGAGTGGGAAATGTTTACCAGAG 2407
Qy 2989 aagaggtcacaaatgagagaaaatcattcacttgccttgcattgaaccatgccacagtg 3048
Db 2408 AGGAGGTTATAATAGAGAGATTTACTTACTTACTTGTGATTTGATTTGATTTGATTTGATTTG 2467
Qy 3049 ccattgccaaaaatttaccctgagaatgaaataatgaaactggtgaaactgagaaatgtcataa 3108


```
Qy 5245 actatttacttggggttaactccgctgaggaactgcataccaaattctgcgaaagagtt 5304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4688 ACTATTTACTTGGAGTAGCTCCGCGCTCAAGAACTGTTTGCCAAATTTCTGCTGAAGGAGAA 4747

Qy 5305 acagtgaactctgtagctatctattcttactacacaaatttctctctctccaggaaat 5364
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4748 TCAGTGCACCTGTGTAGTATCTATCTTTTACCCACAAATTTGTTCTGCTTTTCCAGGAAT 4807

Qy 5365 attgggactgttaagcccttctccagaggtggtgagatccctgccttactactaaat 5424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4808 ATTGGGATACATAAGGCCCTTACTACAGAGGTGGTGTGGAGATCCCTGCGCTTACTCAAGT 4867

Qy 5425 gtttgaagcaaaaaacacccgtgtgtcaggtaccctagaataaaagaaatagtttggatagagc 5484
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4868 CTTTGAAGCAGAAAAGTGCTGTGTGTCAGGTACCTCAGAAAAGAAATAGTTTGATAGAGC 4927

Qy 5485 ttctgtatgactatagctgctctctgaatcaagcttctctatttcagggtgccacggtctg 5544
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4928 TTCTTGAGGACTACAGCTGTCTTTCTAAATCAGGCTTCTCATTAGGTGTCCACGGTCTG 4987

Qy 5545 cagatgatgacgaagacatccctgtccctgtcccttttctgtgggctatactattgtctc 5604
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4988 CAGATGATGACGAACATCTCTGTCTCTCTTTTCTGTGGGGCCATCCTGTGTCTC 5047

Qy 5605 agaacatttctgcagagaaaattgtgaacgggaaggttggagcttgcatttttccagc 5664
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5048 AGAACATCTGTGCCAAGAANAATAGTGAATGGGAAGAGGTGGAGCGTGCCTTTTTCATG 5107

Qy 5665 cacttcaactgtgagccggagttctgctatttctctaaataacagagaatgccgagtgctcc 5724
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5108 CGCTTCACTTGTGTGTGGAGTGTGCTCTCTCTCTTTTCTGTGGGGCCATCCTGTGTCTC 5167

Qy 5725 tggttgaagtaaaagccagaggtgtgctctatccagctccttacttgatgaaatagag 5784
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5168 TGGTGAAGGAAAAGCCAGAGGCTGTGCTCTACCCAGCCCTTACTTGGATGAATATGGAG 5227

Qy 5785 aaacagaccctggcctgaagagggggaaccccttctattctcgtgagcgggtatcggga 5844
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5228 AAACAGACCCAGGCTAAACAGAGGAAACCCACTTCATTTATCTCGGAGCGGTATCGGA 5287

Qy 5845 agctccatttggctggcaacaactgcattatagagagattgctagagcgaagaga 5904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5288 AGCTGCATTTGGTGTGCAACAGCAGCTGCATTTATAGAAGAGATTTGCTCGGAGCCAGGAGA 5347

Qy 5905 ctaatcagatgttatttggattcaactggcaggttactgtgagctccaactgtgctcaag 5964
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5348 CTAATCAGATGCTATTTGGATTTAACCTGGCAGTTTACTCTGAGCTTCAAGTCTGCCTCAAG 5407

Qy 5965 acaatcacaatgacgacagtagtaaaagcgtgattcaaaaaattatggaaaaactttctgagg 6024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5408 ACATCATGAGTGACATCAATAATAAGACTGATCTAAAATTTAGAGAACTTTTCTGAGG 5467

Qy 6025 gctgggaagatattggagggtcttttgcctccatgtccaggttccacttacatcaataaaat 6084
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5468 ACGGGGAAGTATTGGAGGCTCTTTTCATCCATCCTCCAGATTTCACACATTAATAAAAT 5527

Qy 6085 atttcttaatggagattgttcttcaattagcaaacatatcttccacgggaaaaa-aggac 6143
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5528 ATTCTTAATGGAAATATTGTTTCAATTATCAACATTAAGCTTCAAGGGAAAAACAAGAC 5587

Qy 6144 atagatcaactctgt- 6160
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5598 ATGATTAATGTTTATGTGTTCTAGAACACTAAAGAAATAGCTTGTTCATCCAAAGTGTCTAT 5647

Qy 6161 atgtgctagatttccagggaatttattcccttctcataatttggctcatttcttatttttt 6220
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5648 TTCTGCTTAATTTCCAGAAAACTCTTTTCCCTTCATTAACCTGCTTCTAGTTCATTTCATAT 5707

Qy 6221 catcctattggtagatgaagtcacgttcaacagttgttagacatttttattgttgggttaac 6280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5708 CACCCACCTGTTTAATGAGTGCATTAATTAAGCATTTGTGGACATTTCTCCCATCTCGCTAAC 5767
```

```
Qy 6281 tcttctgcaattttgtattgtgtgttt 6307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5768 ATCTCTGCACCTTTGTATTGTGTGTTT 5794

RESULT 3
AF061555 6395 bp mRNA ROD 10-JUL-1998
LOCUS Mus musculus ubiquitin-protein ligase E3-alpha (Ubr1) mRNA,
DEFINITION complete cds.
ACCESSION AF061555
VERSION AF061555.1 GI:3170886
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 6395)
AUTHORS Kwon,Y.T., Reiss,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K.,
Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.
TITLE The mouse and human genes encoding the recognition component of the
N-end rule pathway
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (14), 7898-7903 (1998)
MEDLINE 98318583
REFERENCE 2 (bases 1 to 6395)
AUTHORS Kwon,Y.T. and Varshavsky,A.
TITLE Direct Submission
JOURNAL Submitted (24-APR-1998) Division of Biology, 147-75, California
Institute of Technology, 1200 E. California Bl., Pasadena, CA
91125, USA

FEATURES
Location/Qualifiers
1..6395
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/cell_line="MEL-C19"
1..6395
/gene="Ubr1"
1..114
/gene="Ubr1"
115..5388
/gene="Ubr1"
/note="UBR1p; recognition component of the N-end rule
pathway"
/codon_start=1
/product="ubiquitin-protein ligase E3-alpha"
/protein_id="AAC40165.1"
/db_xref="GI:3170887"
/translation="MADEMDCAERMDVSPPEPLAPQRPASWMDQVDFYTFALHHLA
OLVPEIYFAEMDPDLEKQESVQMSILTPLEWYLFGEDDPDICLEKXKHSAGATQLCGKV
FKSETTYSCRDCAIDPTCVLCMDFQSSVHKHRYKMHSTGGGCDGCDTEAKVTG
PCVDHEPRAGTTKESLHCLNEEVIAQARRIFPSVIKIYVEMTIWEKEKELPPELQ
ITRENNRYCVLFNDEHSHYDHVYSLQRALDCELAELHTTAIDKEGRRAVKAGVY
ATCQEKEDIKSHSENVQHPHLVHLVSMVAHQKALRLGSSWMNKIMSYSDERQI
PCQALFVEEPGSENPCLISRLMLDAKYLKGARKILHELIFSFFMEMEYKKLFAMEF
VKYKQLOKEVYISDDHRSISITALSVMQLTPTLARHLIEBQNVISVTETLLEVL
EYLDNRNKFNGQYSQDLGRVYAVICDKYILISKPTWTERLRAQLLEGPSFLKI
LTCMGMEIIRRVQGHIEVDPDWEAAATAIQMLKNLLMFQEWCAEDBLLLVATKE
CHKAVMRCSTNFMSTKTIVQCLGSHLETSKYKVEDLVSIHLPSRLTLAGLHVLRSR
LGASRLHEFPVDFDSQVEVLVPLRGIVLVAQVVAEMRRNGSLISQVYQDVVK
CREMYDKDIIMLOIGASIMDPNKFLLLVORYELTDAFNKTISKQDOLIKOYNFLI
EEMQVLIYIGERYVPGVGNVREEVIMREITHLICIEPMPSAIALRNLENENNET
GLENVINKVATFKKPGVGHGVELKSLDKDFNMFFHYHYSKTQHSKAMOKKRKK
ENKDEALPPPPPEFCFAPFVKVNNLLSCDVIYILRTIFERAVDTSESNLWTGMLQMA
FHIALGLLEEKQQLQKAPEEVAFDFYHKASRLGSSMAQNIOMLLERLKGIPQLE
GOKDMITWILQWFDTVKRLREKSLWATTSGLGCIKSEETIDHKEMAEKRRKAEAR
LHRQKIMQMSALQKNFTETHKLMYDNTSEVTKEDSIIMEESTSAVSEARLALGPK
RGPATVEKVLTCILQEBQEVKLENNAMVLSACVQKSTALTQHRCKPVDHLGETLDP
LFMPDPLADLHGTTCGCGHVMHVCWKYFEAVOLSSQQRHVDLDESIEGJCLPCK
SLCNTVIDIIPLOPKRINSNAALQLTLARWTQTLVARIISGYNIKHAKGEAPVP
VLFNQMGDPSTFEFHSILSFQSSVKYSNISKEMVILFATYIRIGLVKVPDELDP
VPMWTWSTCAFTIOAIENLGDGPKPLFGALQNRQHSGLKALMQFAVQARATCPQVLI
HKHLARLLSVILPNLQSENTPLLQSVDLFHLVGVAVLAFPSLYWDDTVDLQPSPLSSS
```


Ds 4088 CTTGGAGCACGTGCTGCTTCCACCATCCAGGCAATCGAAAAACCTGTTGGGAGATGAAGGAA 4147
QY 4717 aacctctgttggagcaacttcaaaaatagcagcacaataatgctgtaagaacataatgcagt 4776
Ds 4148 AACCTTATTTTGGAGCACTTCAAAATAGACAGCATACGCGTCTGAAGCGCTAATGCAGT 4207
QY 4777 ttgcagttcacagagattacctgtccctcaggtcctgatacagaacaacatctggttcgc 4836
Ds 4208 TTGCAGTTGCACAGAGGCTTACCTGCGCTCCTCAGGTCTCTGATACACAACATCTGGCTCGGC 4267
QY 4837 ttctatcagttgttcttctcaataaaaaatcagaagatacacacatgcctctctctatag 4896
Ds 4268 TCCTGTCACTTATCTTCTTAACCTGCAATCACAANAATACACAGGCCCTCTGTCTGTGG 4327
QY 4897 atctgtttcatgttttggggggtgctgtgttagcattcccatctgtattggagatgcc 4956
Ds 4328 ATCTCTTCCATGTCTGTGTCGGCGCAGTCTTAGCGTTCCCATCTCTGTATTTGGGATGACA 4387
QY 4957 ctgttgatctgaagcctctctcagtttagttcttctcctataacacacettatctcttcatt 5016
Ds 4388 CCGTGGATCTGCAGCGCTGCGCCACTTAGTTCTTCATATAACCCACCTCTATCTCTTCCATC 4447
QY 5017 tgatcacattggcacacatgcttcagatactactactacagtagaca-----cag 5064
Ds 4448 TGATCACCATGGCGCATGCTTCAGATACTCTCTACACAGATACAGATCTGTCTCCAG 4507
QY 5065 gctaccccttctcaggttccaagaagacagtgaaagaggtcattccgcatctctttct 5124
Ds 4508 GCGCGCGCTTGTGAGGGTGAAGAGGATAGTGAGGAGGCTCGCTGTGCATCTGCTTCT 4567
QY 5125 ttgcagaattctcgaataataaagtggtccatgggtgtgatatctcctgctgtgatt 5184
Ds 4568 TTGTGAAGTGTGCGAGCACAGACAGCGGCTCACTGGGTGCGGTGCTCGCGGTGTACC 4627
QY 5185 tgtgggtctcaactgaagaatggcatcaccccttattctgctgtgctgaatttttcc 5244
Ds 4628 TGTGGCTCTCCCTGAGAGCGGATCACCCCTTACCTCCGCTGTGCTGCACTCTTTCC 4687
QY 5245 actattacttgggttaactccgctgaggaactgcataccaattctgcagaagagagt 5304
Ds 4688 ACTATTTACTTTGAGTAGTCCGCCCTGAAGAACTGTTTGCCAATTTCTGCTGAAGGAGAT 4747
QY 5305 caagtcaactctgtagctatctatttaccatacaaatgttccctgctcttccaggaaat 5364
Ds 4748 TCAGTGCATCTGTAGCTATCTATCTTTTACCACCAAAATTTGTTCTGCTTTTCCAGGAAT 4807
QY 5365 attggataactgaagccctgtctccagagtggtggtgcagatcctgcctactaaact 5424
Ds 4808 ATTGGATACCATTAAGGCCCTTACTACAGAGGTGGTGGAGATCTGCGCTTACTCAAGT 4867
QY 5425 gtttgaagcaaaaaacacccgtgtgaggtacccttagaagaaagaaatagttttagatagc 5484
Ds 4868 CTTTGAAGCAGAAAGTGTGTGTGTCAGGTACCCCTAGAAAAAGAAATAGTTTGCATAGAGC 4927
QY 5485 ttctgatgactatagctcctcctgaataaagcttctcatttcagggtgcccaggtctg 5544
Ds 4928 TTCTCTGAGGACTACAGCTGTCTTCTAAATCAGGCTTCTCAGTTAGGTGTGCCAGGTCTG 4987
QY 5545 cagatgatagcaagacatctgtcctcctgcctcttctgtggggtactactatgttctc 5604
Ds 4988 CAGATGATGAGCAAGACATCT 5047
QY 5605 agaacaatttgcctccaggaattgtgaacgggaagaggttgagcttgcatttttcaag 5664
Ds 5048 AGACATCTGTTCACCAAGAAATAGTGAATGGGGAAGAGGTGGAGCGTTCGTTTTCATG 5107
QY 5665 cacttcattctggagcggagctgctgacttcttcttaaaaaatcagagaatgccagagtggtcc 5724
Ds 5108 CGCTTCACTTGTGTGGAGTCTGCATTTCTCTAAATCCAGAGAATGCAGGCTGGTCC 5167
QY 5725 tgggtgaaggttaaacgagaggtgtgctctatccagctccttacttggatgaatatggag 5784
Ds 5168 TGGTGAAGGAAAAAGCCAGAGGCTGTGCCTACCCAGCCCTTACTTTGGATGAATATGGAG 5227

RESULT 4

AK027803
LOCUS Homo sapiens CDNA FLJ14897 fis, clone PLACE1004743, weakly similar to PROBABLE N-END-RECOGNIZING PROTEIN. PRI 15-MAY-2001
DEFINITION AK027803 2550 bp mRNA
ACCESSION AK027803 GI:14042751
VERSION AK027803.1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
clone:PLACE1004743.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
Wakunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Makamata,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Sasaki,N.
NEDO human cDNA sequencing project
TITLE Unpublished
JOURNAL 2 (bases 1 to 2550)
REFERENCE Isogai,T. and Otsuki,T.
AUTHORS Direct Submission
TITLE Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
JOURNAL NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction;
5'- & 3'-end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)

and Department of Virology, Institute of Medical Science,
- University of Tokyo.

FEATURES

Location/Qualifiers

source

1. .2550

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="PLACE1004743"

/tissue_type="placenta"

/clone_lib="PLACE1"

/note="cloning vector: pME18SFL3"

117. .>2550

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB55380.1"

/db_xref="GI:14042752"

/translation="MTIWEEKEKLPELOIREKNERYCYVLFNDEHHSYDHVITYSLOR
ALDECELAQAQLHTTAIDKERRAVKAGAYACQAEKEDIKSHSENYSORPLHVEVLHS
EIMAHQKFLRGLKSNWKIMSYSDRFQIFCQACLRPEEDSNPCLISRLMLDAKLY
KGARKTLHLIFESSFMEYKKLFAMEFYKKQLOKEYISDDHDSISITIALSVOM
FTVPTLARHLIEFONVISVITETLLEVLPEYLDNRNKNFQOYSQDKLGRVAVICDL
KYLISKPTIWTERLMOFLEGRFSFLKILTCMOGMEIEIRROVGOHLEVPDPEAATA
IQMLKNILLMQEWCACDELLLVAYKECHKAVMRCSTSFISSSKITVWVSCGHSLET
KSTRVEDIVSLHPLUSRLAGLHVRSLRGAVSRHFEVFSFEDQVEVLVEYPLRCL
VLVAQVADMMRRNGLSLISQVEFYQDVCKREEMDKDIIIMLQIGASLMDPNKFLFLY
LQRYELAEAFNKTISTKQDOLIKOYNTLIEMLQVLIYIVGERYVPGVGNVKEEVTM
RETILHULCTEPMPSHAIKLNLENENNETGLENVINKVATFKKPGVSGHGVLELKDES
LKDNNYFVHYSKTOHSAKPHMOKKRRKOENKDEALPPPPPECFAPAFKVINLLNCD
IMMYILRTVEERAIIDTSLNLTWTEGMLQMAFHILALGLEKQQLQKAPESBEVTFDFH
KASRLGSSAMNOMLEKLUKIPQLEGQDKDMITWIIQMFDTVRLRREKSLIVATTSG
SESINKDEITHDK"

BASE COUNT

830 a 454 c 558 g 708 t

ORIGIN

Query Match 40.4%; Score 2546.8; DB 9; Length 2550;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2548; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1189 atcatgaacctggaagacaggttactataaaagagaattcaacgtgtccgttgatgaag 1248

Db 1 ATCATGAACCTGGAAGACGAGGTACTATAAAGAGAATTACGCTGTCGCTGGAATGAAG 60

Qy 1249 aggtlaattgtccagcaggaataatttcttcagtgataaaatatgtcgtgataatga 1308

Db 61 AGGTAAATGTCCAGCCAGCAAAATATTCTTCAGTGATAAATATGCCGTAGAAATGA 120

Qy 1309 ctaataaggaaggaagaaactgcctcctgaactccagataagggagagaaatgaaa 1368

Db 121 CTATATGGGAAGAGAAAAAGAACTGCCTCTGAACCTCCAGATAAGGCAGAAAAATGAAA 180

Qy 1369 gatactattgtctcttttcaatgatgaacacacattcatatgaccacgtcatatagacc 1428

Db 181 GATACTATTGTGCTCTTTCAATGATGAACACCATTCATATGACCACGTCTATATACAGCC 240

Qy 1429 tacaaagagctcttgactgtagctgcagagccagtcagtcaccactgccattgaca 1488

Db 241 TACAAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTCATACCACCTGCCATTGACA 300

Qy 1489 aagaggtctcggtgtgttaaagcggagcttatgctgtctgcaggaagcaaggaag 1548

Db 301 AAGAGGTCGTCGGCTGTTTAAAGCGGAGCTTATGCTGCTGCGCAGGAAGGAAG 360

Qy 1549 atataaagagtcattcagaaaaatgtctctcaacatccactcatgtagaagtattacact 1608

Db 361 ATATPAAAGAGCTCATTCAGAAAAATGTCTCTCAACATCCACTTCATGTAGAAGTATTACACT 420

Qy 1609 cagagattatgctcatcagaaatttgccttgcctgtctgttgcctggaataacaaaaatta 1668

Db 421 CAGAGATTATGGCTCATPCAGAAATTTGCTTTGGCTCTTTGGTTCTCTGGATGAACAAAAATTA 480

Qy 1669 tgagctattaagtgacttttaggcagatcttttgcacagcatgccttagagaagaaacctg 1728

Db 481 TGACCTATTCAAGTGACTTTTAGGCAGATCTTTTGCCCAAGCATGCTTTAGAGAAACCTTG 540

Qy	1729	actcggagaaatccctgtctcataaagcaggttaatgctttgggagtgcaaaactttataaag	1788
Db	541	ACTCGGAGAATCCCTGTCTCATAAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAG	600
Qy	1789	gtgccgtaagatcccttcgaattgatcttcagcaggtttttttatggagatggaataca	1848
Db	601	GTGCCGTAAGATCCCTTCATGAATGATCTTCAGCAGTGTTCATGAGAGTGAATACA	660
Qy	1849	aaaaactctttctatggaatttggaagtattataaaacaactgcagaaagaatatatca	1908
Db	661	AAAAACTCTTTGCTATGGAATTTGTGAAGTATTATATAACAACATGCAAAAGAATATATCA	720
Qy	1909	gtgatgatcatgacagaaagtatctctataactgcacttcagttcagatgtttactgttc	1968
Db	721	GTGATGATCATGACAGAAAGTATCTCTATACAGCACTTTCAGTTCAGATGTTTACTGTTC	780
Qy	1969	ctactctggctgcacatcttattgaagagcagaatgtatctctgtcattactgaaactc	2028
Db	781	CTACTCTGCTCGACATCTTATTGAAGACAGAAATGTTATCTCTGTCTTACTTACTGAAACTC	840
Qy	2029	tctagaagtcttaccctgagttacttggacagagaacataaaattccactctccagggttata	2088
Db	841	TGCTAGAAGTTTACCTGAGTACTTGGACAGGAACAATAAATCAACTCCAGGGTTATA	900
Qy	2089	gccagacaaattggaaagagtatatgcagtaatatgtacacctaaagatatctcctaatca	2148
Db	901	GCAGACAAATTTGGGAAGAGTATATGCAGTAAATATGTGACCTTAAAGTATATCTTGATCA	960
Qy	2149	gcaaaccccaataatggacagaaagattaagaatgcagttcccttgaaggttttctcgtctt	2208
Db	961	GCAACCCACAAATATGGACAGAAAGATTAAGAAATGCAGTTCCTTTGAAGGTTTTTCGATCTT	1020
Qy	2209	ttttgaagattcttacctgtatgcagggaattggaaagaaatccgaagacagaggttgggcaac	2268
Db	1021	TTTTGAAGATTCTTACCTGTATGTCAGGGAATGGAAGAAATCCGAAGACAGAGTTGGGCAAC	1080
Qy	2269	acattgaagtgcattgcgtgaggtgccttgcattctacagatgcattataaagaata	2328
Db	1081	ACATTGAAGTGGATCCTGTGATGGGAGGCTGCCATTGCTATACAGATGCAATTTGAAGAATA	1140
Qy	2329	ttttactcatgttccaagagtggtgtgcttgcagaaagaaactcttacttgcgtgcttata	2388
Db	1141	TTTTACTCATGTTCCAAAGAGTGGTGTGCTGTGATGAAGAACTCTTACTTGTGCTTATA	1200
Qy	2389	aagaatgtcaaaaagctgtgatgaggtgcagttaccagtttcatatctatgtagcaagacag	2448
Db	1201	AAGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	1260
Qy	2449	tagtacaatcgtgtgacatagtttggaaacaaagtcctcagagatctcagagattctg	2508
Db	1261	TAGTACAATCGTGGACATAGTTTGGAAACAAGTCCACAGAGTATCTGAGAGATCTTG	1320
Qy	2509	taagcatatcatctgccaactctcaggaacctgtgctgtgcttcattgacgttttaagcagcc	2568
Db	1321	TAAGCATACATCTGCCACTCTCTAGGACCCCTTCTGCTTTCATGTACGTTTAAAGCAGGC	1380
Qy	2569	tggtgtgctgtttcaagagtcgatgaattgtgctttttgaggactttcaagtagaggtac	2628
Db	1381	TGGTGTCTCTTTCAAGACTGCATGAATTTGTGTCTTTTGGAGGACTTTTCAAGTAGAGTAC	1440
Qy	2629	tagtggataatcctttacagttctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2688
Db	1441	TAGTGGAAATATCTTTTACCTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1500
Qy	2689	gaagaaatgagctgtctcttattagccaggtgttttattaccagagattgtaagtgcagag	2748
Db	1501	GAAGAAATGGACTGTCCTCTTATTAGCCAGGTGTTTATTACCAGAGATGTTTAAAGTGCAGAG	1560
Qy	2749	aagaaatgatgataaagatacatcatccttcagattgtgtgcatctttaaattgagatccca	2808
Db	1561	AAGAAATGTATGATAAAGATATCATCATGCTTCAGATTGGTGCATCTTTTAAATGATGCCCA	1620

```
QY 2809 ataaagtcttctgttactgttactcagaggtatgaacttgcgaggtcttttaacaaagacca 2868
|||||
Db 1621 ATAAGTCTTGTTACTGTGTTACTTCAGAGGTATGAACCTTGCAGGAGCTTTTAAACAAGACCA 1680
|||||
QY 2869 tatcacaaaaacaccaggtttgattaaacaataataacacataaagaaatgcttc 2928
|||||
Db 1681 TATCTACAAAAGACACAGGATTTGATTAAACAATATAACACTAATAGAGAATGCTTC 1740
|||||
QY 2929 aggtctctatctatattgtgggtgagcgttatgtacctggagtgggaaatgtgacaaaag 2988
|||||
Db 1741 AGGTCTCATCTATATGTGGGTGAGCGTATGTAGCTGGAGTGGGAAATGTGACCAAAG 1800
|||||
QY 2989 aagaggtcacaaatgagagaaatcattcacttgccttctgcatgaacccatgcacacagt 3048
|||||
Db 1801 AGAGGTCAATGATGAGAGAAATCACTTCTGCTTTGTCATTGAACCCATGCCACACAGTG 1860
|||||
QY 3049 ccattgccccaaatttaccctgagaatgaataatgaacactggcttagagaatgtcataa 3108
|||||
Db 1861 CCATTGCCAAAATTTACCTGAGAATGMAAATAATGAACCTGGCTTAGAGAATGTCTATA 1920
|||||
QY 3109 acaaaagtgcacatttaagaacaccgggtgtatcaggcccatggagtttatgaactaaaag 3168
|||||
Db 1921 ACAAAAGTGCCACATTTAGAAACACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG 1980
|||||
QY 3169 atgaatcactgaagaacttcaatatatacttttattcattactccaaaaccacagcataga 3228
|||||
Db 1981 ATGAATCACTGAAGAGCTTCAATATATGCTTTTATCATTTCTCAAAACCCAGCATAGCA 2040
|||||
QY 3229 aggtgaaacatatgcagaagaagaagagaaacaagaagaacaaagatgaagcattgcccgc 3288
|||||
Db 2041 AGGCTGAACATATGCAGAGAAAGAGAGAAACAGAAACAAAGATGAAGCATTCGCCGC 2100
|||||
QY 3289 caccaccacctctcgaattctgcccgtgtttcagcaaaagtgttaacctctcgaactgtg 3348
|||||
Db 2101 CACCACCACTCTGAATTTCTGCCCTGCTTTTTCAGCAAAAGTGATTAACCTTCTCAACTGTG 2160
|||||
QY 3349 atatcatgatfacattctcagagaccgtatttgagcgggcgaatagacacagattctaaact 3408
|||||
Db 2161 ATATCATGATGTACATTTCTCAGGACCGTATTTGAGCGGGCAATAGACACAGATTTCTAACT 2220
|||||
QY 3409 tgtgacqaagggatgctccaaatggcttttcataattcttgccattgggtttactagaag 3468
|||||
Db 2221 TGTGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG 2280
|||||
QY 3469 agaagcaacagcttcaaaaagctcctgaagaagaagtaaacatttgacttttatcataaag 3528
|||||
Db 2281 AGAAGCAACAGCTTCAAAAAGCTCTGGAAGAAGATAACATTTGACTTTTATCATAAGG 2340
|||||
QY 3529 ctccaagattgggaagttcagccatgaatatatacaaaatgcttttggaaaaactcaaaagaa 3588
|||||
Db 2341 CTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAAAGGA 2400
|||||
QY 3589 tccccagttagaagcccaagagacatgataacgtggacttcagatgtttgacacag 3648
|||||
Db 2401 TTCCCCAGTTAGAAGCCAGAGGACATGATTAACGTGGATCTTCAGATGTTTGACACAG 2460
|||||
QY 3649 tgaagcattagaagaaaaatctgttttaattgttagcaaccacatcagatcggaatcta 3708
|||||
Db 2461 TGAAGCGGTTAAGAGAAAAATCTGTTTAAATTTGTAGCAACCATCAGGATCGGAATCTA 2520
|||||
QY 3709 ttaagaatgatgagattactcatgataaag 3738
|||||
Db 2521 TTAAGAATGATGAGATTACTCATGATAAAG 2550
|||||
```

```
RESULT 5
AR030785 LOCUS AR030785 1001 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5861312.
ACCESSION AR030785
VERSION AR030785.1 GI:5943999
KEYWORDS Unknown.
SOURCE
```

```
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian UBRI1
JOURNAL Patent: US 5861312-A 2 19-JAN-1999;
FEATURES Location/Qualifiers
source 1..1001
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN /organism="unknown"
```

Query Match 15.8%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 2.6e-240;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 2759 atgatcccaataaagttcttctgacttgcagaggtatgaacttccgaggtttt 2858
|||||
Db 1 ATGGATCCCAACAAGTCTTGTACTGTGTTCTGAGAGGTATGAACCTGCCGAGGCTTTT 60
|||||
QY 2859 aacaagaccatattctacaaaagaccaggtttgtattaaacaataataacactaaagaa 2918
|||||
Db 61 AACRAAGACCATATCTACAAAAGACCAGGATTTGATTAAACAATATAACACTAATAGAA 120
|||||
QY 2919 gaaatgtctcagggtcctcatctatatattgtgggtgagcgttatgtacctggagtggaat 2978
|||||
Db 121 GAAATGCTTTCAGGTCTCATCTATATTGTGGGTGAGCGTTATGTACTGTGAGTGGAAAT 180
|||||
QY 2979 gtgaccaaaagaggtcacaatgagaaatcattcacttcttgcattgaacccatg 3038
|||||
Db 181 GTGACCAAAAGAGGTGACAAATGAGAAATCATTTCACTTGTGCAATTAACCCATG 240
|||||
QY 3039 ccacacagtgccattgccaaaatttacctgagaatgaaaataatgaacactggtctagag 3098
|||||
Db 241 CCACACAGTGCCATTGCCAAAATTTACTGAGATGAAAATTAATGAACACTGGCTTAGAG 300
|||||
QY 3099 aatgtcataaacaagtggccacatttaagaaaccaggtgtatcaggccatggagttat 3158
|||||
Db 301 AATGTCATAAAACAAGTGCCACATTTAAGAAACCAGGTGTATCAGGCCATGAGGTTTAT 360
|||||
QY 3159 gaactaaagatgaatcaactgaagacttcaatatgtacttttatactactccaaaacc 3218
|||||
Db 361 GAATCAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
|||||
QY 3219 cagcatgaagagctgaacatatgcagaagaaagagaaacaaagaaacaaagatgaa 3278
|||||
Db 421 CAGCATAGCAGGCTGAACATATGCAGAAGAAAAGAGAAAACAAGAAAACAAGATGAA 480
|||||
QY 3279 gcattgcccgcaccaccacctcctgaattctgcccctgcttttcagcaaaagtgttaacctt 3338
|||||
Db 481 GCATTGCCGCCACCACCACCTCTGAAATTTCTGCCCTGCTTTTCAGCAAAAGTGAACCTT 540
|||||
QY 3339 ctcaactgtgatacatgatgtatacttcaggaccgtatttgagcgggcaatagacaca 3398
|||||
Db 541 CTCAACTGTGATATCATGATGATACATTTCTAGGACCGTATTTTGGCGGCAATAACACA 600
|||||
QY 3399 gattcctaactgtgagcgaagggatgctccaaatggcttttcatattctcgcatgggt 3458
|||||
Db 601 GATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTTCATATTTCTGCAATTTGGGT 660
|||||
QY 3459 ttactagaagaagaagcaacagcttcaaaaagctcctgaagaagaagtaacatttgacttt 3518
|||||
Db 661 TTACTAGAAGAGAAGCAACAGCTTCAAAAAGCTCTGGAAGAAGAGTAACATTTGACTTT 720
|||||
QY 3519 tatcataagccttcaagattgggaagttcagccatgaatatatacaaaatgcttttgaaaaa 3578
|||||
Db 721 TATCATAGGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
|||||
QY 3579 ctcaaaagaaattccccagtttagaagggcagaggaacatgataaactgatacttcagatg 3638
|||||
Db 781 CTAAAGGAATTTCCCACTTAGAAGGGCAGGAAGGACATGATAACCTGGTACTTTCAGATG 840
|||||
```


Qy 3639 ttgacacagtggaagcgattgaagaaaaatcttgtttaaatttagcaaacacacatcagga 3698
Db 841 TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTGTGTTAAATTGACCAACACATCAGGA 900
Qy 3699 tcggaatctattaaagaatgaagattactcatgataaagaaaaaacagacagaacaaaaaga 3758
Db 901 TCGGAATCTATTAAAGAAATGATGAGATTACTTCATGATAAGAAAAAGCAGAACGAAAAAGA 960
Qy 3759 aagctgaagctgactgagctacatccgacagaagatcatggc 3799
Db 961 AAAGCTGAAGCTGCTAGGCTTCATGCCAGAGATCATGGC 1001

RESULT 6
AR121464
LOCUS AR121464 1001 bp DNA PAT 16-MAY-2001
DEFINITION Sequence 2 from patent US 6159732.
ACCESSION AR121464
VERSION AR121464.1 GI:14105040
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian Ubr1
JOURNAL Patent: US 6159732-A 2 12-DEC-2000;
FEATURES Location/Qualifiers
source 1..1001
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN

Query Match 15.8%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 2.6e-240;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2799 atgagtcacaataagttctgttacttggtacttcagaggtatgaacttgcgaggtttt 2858
Db 1 ATGATGCCAACAGTCTGTGTACTTGGTACTTCAGAGGTATGAACCTGCCGAGGCTTTT 60
Qy 2859 acaagaccatattctacaaagaccaggattgtattacaaataataactacactaaatagaa 2918
Db 61 AACAGACCATATCTACAAAGACCAGGATTGTATTAACAATATATACACTAATAGAA 120
Qy 2919 gaatgcttcaggctcctcatctatatgttgggtgagcggttatgtacctggagtggaat 2978
Db 121 GAATGCTTCAGGCTCCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAGTGGAAAT 180
Qy 2979 gtgacaaagaaggtcacaatgagagaataattcacttgccttgcttgcatgaaccatg 3038
Db 181 GTGACCAAGAGAGGTGCACATGAGAGAAATCATTCACCTTGTGCTTGCATTGAACCCATG 240
Qy 3039 ccacagatgcccattgccaaaaatttaccgtgagaatgaaataatgaaactggcttagag 3098
Db 241 CCACAGATGCCATTGCCAAAAATTTACCTGAGAAATGAATAATGAACTGGCTTAGAG 300
Qy 3099 aatgcatcaacaaagtgccacatttaagaacccgggtgtatcaggccatggagttat 3158
Db 301 AATGTCATAAACAAAGTGGCCACATTTAAGAAACCAAGGTGTATCAGGCCATGGAGTTAT 360
Qy 3159 gaactaaagatgaatacactgaagactcaatatgtacttttattacttactcctcaaaacc 3218
Db 361 GAATCAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCTCAAAACC 420
Qy 3219 cagcatagcaaggtcgaacatcatgacagaagaaaggagaaaaaagaacaaagaatgaa 3278
Db 421 CAGCATAGCAGGCTGAACATATGTCAGAGAGAAAGAGAGAAACAAAGAAATGAA 480
Qy 3279 gcattgcccccaccaccactcctgtaatttcgcttccttcagcagaagtgattacatt 3338
Db 481 GCATTGCCGCCACCACCCTCTCTGAATCTGCCTTGCCTTTCAGCAAGAGTATTACCTT 540

Qy 3339 ctcaactgtgatatacatgatgtacattctcaggaacgctattttgagcgggcaatagacaca 3398
Db 541 CTCAACTGTGATATCATGATGTACATTCTCAGGACCGTATTTGAGGGGCGCAATAACACA 600
Qy 3399 gattctaactgtggaccgagggtatgctccaaaatggcttttcattatcttgccattgggt 3458
Db 601 GATTCTAACTTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTCATATTCTGCGATTGGGT 660
Qy 3459 ttactagaagaagaacacagcttcaaaaagctcctcgaagaagaagtaacatttgacttt 3518
Db 661 TTACTAGAAGAGAGAACACAGCTTCAAAAAGCTCTCGAAGAGAGTAAACATTGTGACTTT 720
Qy 3519 taccataaggcttcaagattgggaagttcagccatgaataatacaaatgcttttggaaaaa 3578
Db 721 TATCATAGGCTTCAAGATTGGGAAGTTTCAGCCATGATATACAATGCTTTTGGAAAAA 780
Qy 3579 ctcaaaaggaattccccagtttagaagggccagaagagcatgataacgtggatctcagatg 3638
Db 781 CTCAAAGGAATTTCCCACTTAGAAGGCCAGAGGACATGATAACGTGGATACTTTCAGATG 840
Qy 3639 ttgacacagtggaagcgattgaagaaaaatctttaaattgtagcaaccacatcagga 3698
Db 841 TTTGACACAGTCAAGCGATTAAAGAGAAAAATCTTGTTTAAATTGTAGCAACCATCAGGA 900
Qy 3699 tcggaatctattaaagaatgagattactcatgataaagaaaaaacagacagaacaaaaaga 3758
Db 901 TCGGAATCTATTAGAATGATGAGATTACTCATGATAAGAAAAAGCAGACACGAAAAAGA 960
Qy 3759 aaagctgaagctgctagctacatccgacagaagatcatggc 3799
Db 961 AAAGCTGAAGCTGCTAGGCTTCATGCCAGAGATCATGGC 1001

RESULT 7
AF061556 999 bp mRNA PRI 10-JUL-1998
LOCUS Homo sapiens ubiquitin-protein ligase E3-alpha (UBR1) mRNA, partial
DEFINITION cds
ACCESSION AF061556
VERSION AF061556.1 GI:3170888
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 999)
AUTHORS Kwon,Y.T., Reiss,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K., Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.
TITLE The mouse and human genes encoding the recognition component of the N-end rule pathway
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (14), 7898-7903 (1998)
MEDLINE 98318583
REFERENCE 2 (bases 1 to 999)
AUTHORS Kwon,Y.T. and Varshavsky,A.
TITLE Direct Submission
JOURNAL Submitted (24-APR-1998) Division of Biology, 147-75, California Institute of Technology, 1200 E. California Bl., Pasadena, CA 91125, USA

FEATURES
source Location/Qualifiers
1..999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15q15-q21.1"
/chromosome="15"
1..>999
/gene="UBR1"
1..>999
/gene="UBR1"
/note="UBR1p; recognition component of the N-end rule pathway"
/codon_start=1
/product="ubiquitin-protein ligase E3-alpha"

IIQILLTSCTEENGMDQENPPCEESAVLALYKTLHQYTGSALKKEIPSGMHLRSVRA
GIMPFUKCSALFHYLVNGVPSPPDIQVQPSHEFLHLSYLSLNNLLJCLQFENSEIMN
SLIESCRNSEVRYLEGERDAIRYPRESKNLILNPEDYSLLINQANFSQPSKGGDK
SRAPTICLVGSLUCSQSYCCQTELEDVACTAHTYSCGSGVGIPLRVRECOVLFL
AGKTKGCFYSPPLDDYGETDQGLRGNNPLHLCKERFKIKQLKHQHSVTEIGHAQE
ANQTLVIGIDWHL"

BASE COUNT 1832 a 1199 c 1320 g 1807 t
ORIGIN

Query Match 15.6%; Score 981; DB 9; Length 6158;
Best Local Similarity 55.9%; Pred. No. 2.1e-236;
Matches 2157; Conservative 0; Mismatches 1610; Indels 90; Gaps 12;

Qy	2130	ctaaagtatactctgacagcaaacaccacataatgagacagaagaattaaagaatcgacttc	2189
Db	1	CTCAAGTATGTTAATAAGCAAACTCAATGGTCAGATGAGCTGAGCGCAAGTTC	60
Qy	2190	cttgaaggttttcgatctttttgaagattcttaccctgtatgcagggaatggaagaaatc	2249
Db	61	CTAGAAGGGTTGATGCTTTTGAATTTACTAAATGTATGCAGGGAATGGATCCAATT	120
Qy	2250	cgaagacaggttggggaacacattgaagtgcgtactgattgggagggctgccattgctata	2309
Db	121	ACAGCTCAAGTAGGACAACTATTTGAAATGGAAACAGAGTGGGAAGCAGCCTTCACACTA	180
Qy	2310	cagatgaattgaagaataatttactcatgttccaagatgtgtgtctgtgagaaaga	2369
Db	181	CAATGAATTAACACATGTCATTTCAATGATGAGGACTGGTGTCTTCAGATGAAAAA	240
Qy	2370	ctcttaacttggcttataagaatgtccaaaagctgtgatgaggtcagtlaccagatttc	2429
Db	241	GTGTTAATCGAAGCTTACAGAATATGCTCGCTGCTACTGTGATGCAGTGCATGGTGTAT	300
Qy	2430	atatctagtagcaagacagt----agtcacatcgtgtggaacatagtttggaaacaaagtcc	2486
Db	301	ACTGATGGTGAACAGCCAAATCACACTAAGCATTTTGTGGACATTCAGTGGAAACTATCAGA	360
Qy	2487	tacagatctctaggtatctgtaagcatatcatctgccactctctagacccttgcgtgt	2546
Db	361	TACTGTGTTCCCAAGAAAAGTTAGCATTCACCTCCAGTTTCTCGCTTACTTGCAGGT	420
Qy	2547	cttcatctagctttaagcaggtgggtgtgttccaagactgcataagatttgttctttt	2606
Db	421	TTACATCTATTATTAGCAAAAGTGAAGTGCATATAAAATTTCCAGAGCTCTACCTCTA	480
Qy	2607	gaggaacttcaagtagaggtagttagaataatctcttaoigtctgtgtgtgtgtgtgcc	2666
Db	481	AGTGAACCTTAGCCCAACCCATGTTGATAGAACACACCTCTTAGATGTCTTGTGTGCC	540
Qy	2667	caggttctgctgaagatgtgcgaagaatggagctgtctcttattagccaagtgcttat	2726
Db	541	CNAGTACATCCGGAATGTGGAGAAAGATGGGTCTCTCTAGTAACACGATTATTATAC	600
Qy	2727	taccaagattgaagtcagagaagaatgtatgataaagatacatcatcttccagatt	2786
Db	601	TACCAATAATCGAAATCCAGACGTGAGATGTTTGCACAAAGATAGTAGTAATGCTTCAGACA	660
Qy	2787	ggtgcatttaatggtatcccaataaagtcttcttgaatggttaacttcagagatgaactt	2846
Db	661	GGTGTCTCCATGATGGATCCAAATCATTTCTGTATGATCATGCTCAGCGGCTTTGAACCT	720
Qy	2847	gcagagcttttaacaaagaccatctacaaaa-----gaccacg	2885
Db	721	TATCAGATTTTCTAGTACTCAGACTATGGAAAAAGATTTAGTTCTTGAGATTACCCATAAG	780
Qy	2886	gatttgattaaacaataataactaactagaagaatgtctcaggctcctcatctatatt	2945
Db	781	GATGTTGTTGACGAGACAATACTCTTAATAGAAGAAATGCTATACCTCATTAATGCTT	840
Qy	2946	gtgggtgagcgttatgtacctgtgaggtgggaatgtgacaaagaagaggtgcacaatgaga	3005
Db	841	GTTGGAGAGAGATTTAGTCTCTGGAGTTGGACAGGTAAATGCTACAGATGAATCAAGCGA	900

Qy	3006	gaaatcattcacttgccttgcattgaaccatgccacacagtcgcatggccataaaattta	3065
Db	901	GAGATTATCCATCAGTTGAGTATCAAGCCTATGCTCATAGTAATGGTAAAGTCTTTA	960
Qy	3066	cctgagaatgaaaaataatgaaactggcttagaagaatgcataaacaagtgccacattt	3125
Db	961	CCTGAAGATGAGAACAGGAGACTGGCATGGAGTGAATCGAAGCAGTTGCCCATTTT	1020
Qy	3126	aagaacacaggtgatcaggccaatgagttttatgaactaaagaatgaatcactcaaaagc	3185
Db	1021	AGAAAACTGGGATTAAAGACAGCAGGCATGTATGAATGAACAGAGATGTGCCAAGAG	1080
Qy	3186	ttcaatatgtactttttatctactctcaaaacccagcatagcaaggtcgtgacatatacag	3245
Db	1081	TTCAACTTGTATTTCTATCACTTTTCAAGGGCAGAACAGTCCCAAGGAGGAGCGCAA	1140
Qy	3246	aagaaaaggagaaaaaacaagaatgaagcattgcgccaccacacacccctcctgaa	3305
Db	1141	CGAAAAATTGAAAAGACAAAATAGAGAAGATACAGCACTCCACCTCCGGTGTTCCTCCA	1200
Qy	3306	tctgcctgtcttccagcaaaagtattaaaccttctcaactgtgatatacatgatgtacatt	3365
Db	1201	TTTGTGCCCTCTGTGTTCCAAAGCCTGGTTAACATTTTGCAAGTGCATGTGTGTGCATC	1260
Qy	3366	ctcaggacccgtatttgagcgggcaatagacacagattctcaactgtgtgaccgaaggtatg	3425
Db	1261	ATGGGACAAATTTCTGCAATGGGCTGTGGACATAAATGGATATGCCTGGTCAGAGTCCATG	1320
Qy	3426	ctccaaatggcttttcatattctggcattgggttttactagaagaagaacacagcttcaa	3485
Db	1321	CTGCAAAAGGCTGTTACATTTAATTTGGCATGGCACTACAAGAAGAAAAACAACATTTAGAG	1380
Qy	3486	aaagctcctggaag--aagaagtaacatttgaacttttatcataagggttcaagattggga	3542
Db	1381	AATGTCACGGGAAGAGCATGTAGTAACATTTTACCTTCTCAGAAAGATATCAAAACCTGGT	1440
Qy	3543	agttcagccatga-----atacaaatgcttttggaaaaactcaaaaggaattccc	3593
Db	1441	GAAGCCCAAAAATTTCTCTAGCATACTAGCTATGCTGGAAAACACATACAAAATGCTCC	1500
Qy	3594	cagttagaagggccagaaagacatgataaacgtggatacttcagatgtttgcacacagtgaa	3653
Db	1501	TACCTAGAAAGTCCACAAAGACATGATTCGCTGGATATTGAAGACTTTTAAATGCTTTAAA	1560
Qy	3654	cgattaaaggaaaaaactgttttaattgttagcaacacatacaggaatcggaatctattaa	3713
Db	1561	AAAGATGAGGGGAGA-----GTTCCCTTACCAGTCCCTGGCAGAGACAGAAAGCAACATA	1614
Qy	3714	aatgatggatLactcatgataaagaaaaagcagaaagcaaaagaaagctgaaagctgct	3773
Db	1615	ATGGAAAGAGGTTTCAAGGGCAAAAGACAAAGCTGAGAGGAGAGAAAGACAGATTGCC	1674
Qy	3774	aggctacatcgcgcagaagatcatggctcagatgtctgcctttacagaaaaacttcattgaa	3833
Db	1675	AGACTCGCAGAGAAAGATCATGGCTCAGATGCTGCTGAAATGCGACGCGCAITTTATTGAT	1734
Qy	3834	actcataaactcatgtatgacaaatatacatcagaagaatgctcgtgggaagaagattccattatg	3893
Db	1735	GAATAACAAAGAACTTTTCAGCAGACATTTAGAACTGGATGCTCAACCTCTGCTGTTCTT	1794
Qy	3894	gaggaagagagcaccacagcagtcagtgactactctagaattgcttgggtcctaaacgg	3953
Db	1795	GATCA-----TAGCCCTGTGGCTTCAGATATGACACTTACAGCACTGGGTCACACAAA	1848
Qy	3954	ggtccactgttactgaaaaggggtgctgactgctctcttgccttcccaagaagaacagag	4013
Db	1849	ACTCAGGTTTCTTCAACAAAGACAAATTCGTTTACATGTATATTTGTCRAAGAGGAGCAAGAA	1908
Qy	4014	gtgaaaatagaaaaaatgacctggttattatcgcgcctgtgtccagaagaatctaccgctta	4073
Db	1909	GTTAAAGTGAAAGACGGGCAATGGTCTTGCGACCAITTTGTTTCAGAGATCAACTGTATTA	1968

QY 4074 accgacagcaggggaaacccataagaactctcaggagaagcccttagaccacacttttcagt 4133
Db 1969 TCAAAAACCAAGAGTAATTTAATCAAGATCCAGAAAATAAT---GATCCATTAATTCATG 2025
QY 4134 gatccagacttgccataatgaaactatacaggaagctgtggtcatgaaatgaacgcagtg 4193
Db 2026 CACCCGTGATCTGCTGTGGAAACACACACTAGTACGTGTGGGCACATTATGATGCCCAT 2085
QY 4194 tgcctgcagaagtattttgaaagctglacagctgagctctcagcagcgcattcattgttgc 4253
Db 2086 TGTGTGCAAGAGTATTTGATTCGGTTCAAGCTTAAGAAACAGCAAGGCAACAGAGATTA 2145
QY 4254 cttttt-----tgacttgaaagtggagaataatcttttgccctcttttgcataatct 4301
Db 2146 CGCTTACATACAGAGTATGATGTAGAAAACGGAGAAATCTCTTGCCCCCTTTGTGTAATG 2205
QY 4302 ctgtgcaactgtgatcccaattctcttgcaacctcaaaagataaacaagtgagaat 4361
Db 2206 TTGAGTAATAGTGTATTC-----TCGTCTGCTTCCTCCAGAAATATTTTAAACAC 2259
QY 4362 gcagatgctcttgcctaacttttgacctggcacggtggatcacagactgttctggccaga 4421
Db 2260 AGGTTAAATTTTCAGACCAACCAATCTGACTCAGTGCATTAGAACAAATATCTCAGCAA 2319
QY 4422 atatcaggtataataagaacatgctaaggagaaaaacccaattctctttctttaat 4481
Db 2320 ATAAAAGCATTTACAGTCTTTTGAAGAAAGAAAGTACTCTCTTAATAATGCCCTCTACAAG 2379
QY 4482 caaggaatgggagattctacttggagttccattccatcctgagtttggcgttgagct 4541
Db 2380 AATTACAGAAATGTGGATGAATTACAGCTCCCTGGAAGGTTTCAGGCCTGATTTTCGTCT 2439
QY 4542 tggatataattcaaatagcatcaaggaaatgggtattctctcttggccacaacaatttat 4601
Db 2440 AAGATCCCTTATTTCTGAGAGCATAAAGAAATGCTAAACGACATTTTGGAACTGCTACCTAC 2499
QY 4602 agaattgattgaagtgccacctgatgaagggtatcctcagtgccccatgctgaacctgg 4661
Db 2500 AAGGTGGGACTAAAGGTTTCATCCCAATGAAGAGATCCCTCGTGTGCCCAATAATGTGTGG 2559
QY 4662 agcactgcgcttcactatccaggcaattgaaatctattggagatgaaggaaacct 4721
Db 2560 GTAGCTCGCGGTACACCATCCAAAGCATAGAAAGAAATTTTGAAGTGAAGATAAACCA 2619
QY 4722 ctgtttgagcacttcaaataggcagcataaattgctgtgaaagcatttaagcagtttga 4781
Db 2620 TTGTTTGTGCTCTTACCTTGACAGCTGGATGACTGTCTTAGTTCATTGACGAGATTTGCC 2679
QY 4782 gtgcacagagattacctgtcctcaggtcctgtgatagacaaacatctggtcttctta 4841
Db 2680 GCAGCACACTGGACAGTGGCATFCACTTTTCAGTGGTGCAAGAGACATTTTGTAAACTTTT 2739
QY 4842 tcagttgtcttcctaaataaatacagaagatacacatgcctctctgtctatagatctg 4901
Db 2740 GCATCACTGGTCCCTAATGACAGCCATGAGGAATCTCCATGATATATGATATGACATG 2799
QY 4902 ttctatgtttggtgggtgctgtgttagcattcccatcctctgtattgggtagacctgtt 4961
Db 2800 TTTTATTTATTTGGTGGCTTGTGTGCTGATTTCTCGCTTGCACTG----- 2846
QY 4962 gatctgcgcctctctcagtttagttctctctataaccacctttatctctcatttgatc 5021
Db 2847 --TCAGGATTTTTCAGGATTCAGCCCTGGCACTGGAGACCTTCACATTTTCCATCTGGTT 2904
QY 5022 acctggcacactgtctcagataactactacagtagaacagacc---tacccttgcct 5078
Db 2905 ACTATGGCACACATCATACAGATCTTTACTTACTCTCATGTACAGAAAGAAATGGCATGGAT 2964
QY 5079 caggttcaagaagacagtgagaggtctcattccgcctctctctcttcttcgcagaaattct 5138
Db 2965 CAAGAAATATCCCTTGTGAAGAAAGAAATCAGCAGTCTCTGCTTTGTATAAACACTTCAC 3024
QY 5139 caataacaagtggtccattgggtgtgatatctcctggctggtatttgggtctcactg 5198

Db 3025 CAGTATACGGGAAGTGCTTGAAGAAATACCATCCGGCTGGCATCTGTGGAGGAGTGTC 3084
QY 5199 aagaatggcatcaccccttattctcgtgctgtgctgattgttttccactatttacttggg 5258
Db 3085 AGACCTGGAATCATGCTTCTCTGAAGTGTCTGCTTATTTTTCATTACTTAATGGA 3144
QY 5259 gtaactccgcctgagggaactgcatacccaattctgcagaagagagatcacagtctgt 5318
Db 3145 GTTCTTCCCCCACCGACATTCA---AGTCTCTGGAACAAGCCATTTTGAACATTATGT 3201
QY 5319 agctatctatcttacctacaaaatttctcgtctctccaggaataattggataactgta 5378
Db 3202 AGCTATCTTCTCCCTACCAACAACCTCAATTTGCTCTTTTCAAGAAATAGTGAGATAATG 3261
QY 5379 aggccttgcctccagaggtggtgcagactcctgcttactaaactgttgaagcaaaaa 5438
Db 3262 AATTCACGTGATTGAAGTTGGTGGCTTAACAGTGAAGTTAAAGATATCTAGAAGTGAA 3321
QY 5439 aacacccgtgctgaggtaccctagaaaaaagaatagttgtatagacttctctgactat 5498
Db 3322 AGAGATGCTATAAGATATCCAAGAGAAATCTAACAATTAATAAACCTTCCAGAGGATTAC 3381
QY 5499 agctgcctcctgaatcaagcttctcatttcaggtggtccacaggtctgcagatgatgaga 5558
Db 3382 AGCAGCTCATTAATTAAGCATCCAATTTCTCGTGCCCGAAATCAGGTGGTGATAAGAGC 3441
QY 5559 aagcactcctgcctcgtccttctgtggggtctactatctctcagaacaatttgcgc 5618
Db 3442 AGACCCCAACTCTGTGCTTGTGTGGGATCTCTGCTGTGCTCCAGAGTTACTGTCTGC 3501
QY 5619 caggaaatttgaaacggggaaggttgagcttgcatttttcacgcacttccactgtgga 5678
Db 3502 CAGACTGAACCTGGAAGGGAGGATGTAGAGCTGCACAGCTCACACCTACTCTCTGTGCG 3561
QY 5679 gcgggaatctgcatttctcctaaataacagaaatgcagagtggtcctcgttgaaggtaaa 5738
Db 3562 TCTGGAGTGGGCATCTTCTCTGAGAGTACGGAAATGTCAGGTGCTATTTTAGCTGGCAAA 3621
QY 5739 gcagagagctgtgctctccagctccttacttctggtgatgaatggaagaaacagacctggc 5798
Db 3622 ACCAAGGCTGTTTTTATCTCTCTCTTACCTTGATGACTATGGGAGACCCAGCAGGA 3681
QY 5799 ctgaagaggggcaaccccttctattctcgtgagcgggtatcggaagctccatttggct 5858
Db 3682 CTCAGACGGGAAATCTCTTACATTTATGCAAGAGCGATTCAAGAAGATTCAAGAAGCTC 3741
QY 5859 tggcaacaacactgcattatagaagagattgtcaggagcccaagagactaatcagatgta 5918
Db 3742 TGGCACCACACACAGTGTACAGAGGAAATGGACATGCACAGGAAGCAATCAGACACTG 3801
QY 5919 ttggattccaactggca 5935
Db 3802 GTTGGCATTCAGCTGGCA 3818

RESULT 9

HSX521.
LOCUS HSX521 818 bp DNA PRI 04-OCT-1995
DEFINITION Human partial cDNA sequence, clone x521;.
ACCESSION 247040
VERSION 247040.1 GI:1008975
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 818)
AUTHORS Chianilikulchai,N., Pasturaud,P., Richard,I., Auffray,C. and Beckmann,J.S.
TITLE cDNA selection in the LGMD2A region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 818)

AUTHORS Genethon.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-1994) Genethon, B.P. 60, 91002 Evry Cedex France.
REFERENCE E-mail: Nucleo@genethon.fr
AUTHORS 3 (bases 1 to 818)
 Chiannikulkhai,N., Pasturaud,P., Richard,I., Auffray,C. and Beckmann,J.S.
TITLE A primary expression map of the chromosome 15q15 region containing the recessive form of limb-girdle muscular dystrophy (LGMD2A) gene
JOURNAL Hum. Mol. Genet. 4 (4), 717-725 (1995)
MEDLINE 95359980
FEATURES Location/Qualifiers
 source I. .818
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /clone_lib="CEPH YAC clone 774G4 was used for cDNA selection"
 /map="15q15"
 /tissue_type="fetal muscles"
 BASE COUNT 225 a 177 c 195 g 221 t
 ORIGIN

 Query Match 12.1%; Score 764.2; DB 9; Length 818;
 Best Local Similarity 98.2%; Pred. No. 9.2e-182;
 Matches 783; Conservative 0; Mismatches 13; Indels 1; Gaps 3;

 QY 5302 agtaagtcgaactctgtagctatctctatcttttaactacaattgttctcgtctcttcagg 5361
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 3 ATTCAGTGCACCTCTGTGGTATCTATCTTTACCTACAAATGGGTTCCTGCTTCCAGG 62

 QY 5362 aatatgggatactgtaagcccttgcctcagaggtggtgcagatcctgccttactaa 5421
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 63 AATATGGGATACTGTATAGGCCCTTGCTCCAGAGGTGGTGTGCAGATCCTGCCATTACCAA 122

 QY 5422 actgttgaagcaaaaaaacacgcgtggtcggtaccctagaaaaagaataagttgatag 5481
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 123 ACTGTTGAAGCAAAAAAACACCGTGTGTCAGGTACCTGAGAAAAAGAAATAGTTGATAG 182

 QY 5482 agttctctgatgactatagctgctcctcctgaatcaagcttctcatttcaggctgcccaaggt 5541
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 183 AGCTTCTTGATGACTATAGCTGCTCTCTGAATCAAGCTTCTCATTTGAGGTGCCACGGT 242

 QY 5542 ctgcagatgatgacgaagaacacctgtcctctcctgccttttctgtggagctatactatgt 5601
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 243 CCGCAGATGATGACGAAGAATCCTCTGCTCTGCCCTTTCTGTGGGGCTATACATATGTT 302

 QY 5602 ctcaaacatttgtctccaggaaattgtgaacgggggaaggttggagcttgcatttttc 5661
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 303 CTCAGAACATCTCTGCCAGGAAATGTGAACGGGGAAGAGTTGGAGCTTGCAATTTTTC 362

 QY 5662 acgaacttcactgtgagccggagctgtgcaatttctctaaaaatcagagaatgccagatgg 5721
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 363 ACGCACTTCACCTGTGAGCGGGAGTGTGCAATTTCTTAAATAACAGAAATGCCGAGTGG 422

 QY 5722 tctctgttgaaggttaagccagaggtgtgcctatccagctccttacttggatgaatatg 5781
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 423 TCCTGGTTGAAGGTAAAGCCAGAGGCTGTGCCCTATCCAGCTCCTTACTTGGATGAATATG 482

 QY 5782 gaaaacagaccctggcctgaagagggg - caacccccttcatattatctctgtagcggtat 5840
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 483 GAGAAACAGACCCCTGGCCTGAAGAGGGGCCAACCCCTTCATTTATCTCGTGAGCGGTAT 542

 QY 5841 cggaaagctcatttggtctgtggcaaacacactgcattatagaagagattgtctaggagccaa 5900
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 543 CGGAAGCTCCATTTGGTCTGTGGCAACAACACTGCAATTAAGAAGAGGTGCTTAGGAGCCAA 602

 QY 5901 gagaactaatcagatgttatttggattcaactggcagttactgtgagctccaactctgcct 5960
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 603 GAGACTAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCTGCCT 662

 QY 5961 caagacaatcaaaaatgacagacagtagtaagaggtgattcaaaaattatggaaaactttct 6020

Db	663	CAAGACAATTCACAAATTCAGCACAGTAGTAAGGCTGATTCCAAATTCATGGAATTCCTTCT	722
Qy	6021	gaggctgggaagtattgagaggtcttttgcctccatgctcaggttcacttacatcaata	6080
Db	723	GAGGCTGCTGAAGTATTGGAGGCTCTTTGCTCCATGCTCAGGTTCACTTACATCAATA	782
Qy	6081	aaatattcttaatgga	6097
Db	783	AAATATTTCTTAATGCA	799
RESULT 10			
LOCUS	AK026998	2958 bp	mRNA PRI 29-SEP-2000
DEFINITION	Homo sapiens	CDNA: FLJ23345 fis, clone HEP13679.	
ACCESSION	AK026998		
VERSION	AK026998.1	GI:10439998	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens	hepatoma cell_line:HepG2	cdNA to mRNA, clone_lib:HEBP
clone:HEP13679.			
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (sites)	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,		
	Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,		
	Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	unpublished (2000)		
REFERENCE	2 (bases 1 to 2958)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,		
	Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio		
	Sugano, Institute of Medical Science, University of Tokyo,		
	Laboratory of Genome Structure Analysis, Human Genome Center;		
	Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan		
	(E-mail:cdnal@ims-u-tokyo.ac.jp, Tel:81-3-5449-5286,		
	Fax:81-3-5449-5415)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of		
	International Trade and Industry of Japan; cDNA full insert		
	sequencing; Research Association for Biotechnology; cDNA library		
	construction, 5'- and 3'-end one pass sequencing; Department of		
	Virology and Human Genome Center, Institute of Medical Science,		
	University of Tokyo (partly supported by Science and Technology		
	Agency).		
FEATURES	Location/Qualifiers		
source	1. .2958		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="HepG2"		
	/cell_type="hepatoma"		
	/clone="HEP13679"		
	/clone_lib="HEP"		
	/note="cloning vector pME18SFL3"		
BASE COUNT	940 a 600 c 638 g 780 t		
ORIGIN			
Query Match	10.3%;	Score 651.6;	DB 9; Length 2958;
Best Local Similarity	54.6%;	Pred. No. 2.8e-153;	
Matches 1555;	Conservative 0;	Mismatches 1229;	Indels 66; Gaps 1
Qy	3113	agtggccacatttaagaaacagggtgatcaggccatggagtttatgaactaaagatga	3172
Db	9	AGTTGCCCATTTCAAGAAACCTGGATTAAACAGGACGAGGCATGTATGAACCTGAAACCGA	68
Qy	3173	atcactgaagacttcaatgatgacttttatcattcattcctaaaaaccagcatgcaagc	3232
Db	69	ATGTGCCAAGAGTTCACCTGATTATTCTATCTATCTTTTCAGGCGAGAACATCCGACGC	128
Qy	3233	taacatatgcacacaaagagagaaacaaagaaacaaacataaagcattccgccacc	3292

Qy	5426	tttgaagcaaaaaaacaccctggttcaggtagcccttagaaaaaatagtttgatagact	5485
Db	2310	TCTAAGAGGTGAAGAGATGCTATAAGATATCCCAAGAAATCTAACAAATTAATAAACCT	2369
Qy	5486	tccgtgactatagctgcctcctgaatcaagcttctcatitcagggtgcccaaggtctgc	5545
Db	2370	TCCAGAGATTACAGCAGCCTCATTAATCAAGCATCCAATTTCTCGTCCCGAAATCAGG	2429
Qy	5546	agatgatgagcgaagcatcctgtccctctgctcttctgtggggtctactatgttctca	5605
Db	2430	TGGTGATAAGAGCAGAGCCCAACTCTGTGCTTGTGTGCGGATCTCTGCTGCTCCCA	2489
Qy	5606	gaacatttgtgcocagaaaatttgaaocggggaaagaggttgagagcttcttcaagc	5665
Db	2490	GAGTTACTTGCTGCCAGACTGAACCTGGAAGGGAGGATGTAGGAGCTGCACAGCTCACAC	2549
Qy	5666	acttcactgtggagccgagctcagctatttctctaaaaatcagaagaatccgagtgctct	5725
Db	2550	CTACTCTGTGGCTCTGGAGTGGGCATCTTCTCAGAGTACGGGAATGTCAGGTGCTATT	2609
Qy	5726	ggttgaaggttaaacgagagctgtgcctctccagctcttacttggatgaataggaga	5785
Db	2610	TTTAGCTGGCAAAACCAAGCGCTGTTTATTCTCTCTTACCTTGATGACTATGGGA	2669
Qy	5786	acagaaccttgctcgtgaaggggcaaccccttcattatctcgtgagcggatcggaa	5845
Db	2670	GACCCAGCAGGGATCTACAGCGGGAAATCCCTTTACATTTATGCAAGAGCGATTCAAGAA	2729
Qy	5846	gctccatttgctcgtgcacaaacactcattatagaagagattcctagagccaaagac	5905
Db	2730	GATTCAGAAGCTCTGGCCACCAACACAGTGTCCAGAGGAAATGGACATGCACAGGAAGC	2789
Qy	5906	taatcagatgtatttgattgattcaactggca	5935
Db	2790	CAATCAGACACTGTTGGCTTGACTGGCA	2819
RESULT	11		
AC090514/c			
LOCUS	AC090514	166518 bp	DNA
DEFINITION	Homo sapiens chromosome 15 clone RP11-2016 map 15q15, complete sequence.		
ACCESSION	AC090514		PRI
VERSION	AC090514.1	GI:13129421	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 166518)		
	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.		
TITLE	Sequencing of human chromosome 15 D15S146-D15S117 region		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 166518)		
AUTHORS	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-FEB-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA		
REFERENCE	3 (bases 1 to 166518)		
AUTHORS	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAR-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA		
COMMENT	----- Genome Center		

Center: Multimegabase Sequencing Center
Center code: UMWSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
Drafting center: WIBR
----- Summary Statistics -----
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

FEATURES	Location/Qualifiers
source	1. .166518

```

1. .166518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="Rp11-2016"
/clone_lib="RpCI human BAC library 11"
/note="This clone overlaps Rp11-50M11 AC068727 and
Rp11-473C18 AC068724"
misc_feature
1. .11505
/note="overlap with Rp11-90M11 AC068727"
unsure
47150. .47190
/note="low quality data"
unsure
106560. .106790
/note="low quality data"
unsure
109345. .109360
/note="low quality data"
unsure
114180. .114240
/note="low quality data"
misc_feature
156815. .166518
/note="overlap with Rp11-473C18, AC068724"
BASE COUNT      51718 a 34493 c 32351 g 47956 t
ORIGIN

```

[illegible]

Qy	6282	ctctgcgaattttgtatttgggtctttt	6308
Db	41342	CTTCTGCAATTTTGTATTGGTGTCTTT	41316
RESULT 12			
AK026948	2512 bp	mrna	29-SEP-2000
LOCUS	Homo sapiens cDNA: FLJ23295 fis, clone HEP10637.	PRI	
DEFINITION	AK026948		
ACCESSION	AK026948.1	GI:10439928	
VERSION	AK026948.1	GI:10439928	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens hepatoma cell_line: HepG2 cDNA to mRNA, clone_lib: HEP		
	clone: HEP10637.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (sites)		
	Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,		
	Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,		
	Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
REFERENCE	2 (bases 1 to 2512)		
AUTHORS	Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,		
	Shibahara, T., Tanaka, T. and Nakamura, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio		
	Sugano, Institute of Medical Science, University of Tokyo,		
	Laboratory of Genome Structure Analysis, Human Genome Center;		
	Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan		
	(E-mail: cdnalm@ms.u-tokyo.ac.jp, Tel: 81-3-5449-5286,		
	Fax: 81-3-5449-5416)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of		
	International Trade and Industry of Japan; cDNA full insert		
	sequencing; Research Association for Biotechnology; cDNA library		
	construction, 5'- & 3'-end one pass sequencing; Department of		
	Virology and Human Genome Center, Institute of Medical Science,		
	University of Tokyo (partly supported by Science and Technology		
	Agency).		
FEATURES	Location/Qualifiers		
source	1..2512		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="HepG2"		
	/cell_type="hepatoma"		
	/clone="HEP10637"		
	/clone_lib="HEP"		
	/note="cloning vector pME18SFL3"		
BASE COUNT	785 a 511 c 543 g 673 t		
ORIGIN			
Query Match	7.7%;	Score 487.8;	DB 9; Length 2512;
Best Local Similarity	54.1%;	Pred. No. 6.1e-112;	
Matches 1220;	Conservative	0; Mismatches 987;	Indels 50; Gaps
Qy	3694	caggatcggaattcttaagaatgagatctactcatgataaagaaagacagacgaa	3753
Db	150	CACAGACAGGAAGAACCATTAATGGAGAGAGTTCAAGGGACAAAGACAAGCTGAGAGGA	209
Qy	3754	aaagaaaagctgaagcttagctacatcgccagaagatcatggtcagatgctgcct	3813
Db	210	AGAGAAACGACAGATTGCCAGACTCGCGACAGAGAAAGATCATGGCTCAGATGCTGAA	269
Qy	3914	tacagaaaacttcattgaaactcataaactcatgtatgacaatacatcagaaatgcctg	3873
Db	270	TGCAGCGGCATTTTATTGTATGAAAAACAAAGAACTCTTTTCAGCAGACATATAGAATGGATG	329
Qy	3874	ggaaaagaagattcattatggaggaaagagacacccacagctcagtgactactctagaa	3933
Db	330	CCTCAACCTCTGCTGTTCTTGATCA-----TAGCCCTTGCTTCAGATATGACACTTA	383


```
Db 1438 TTCACATTTTCCATCTCGTGTACTATGCGACACATCATACAGATCTTACTTACCTCATGTA 1497
Qy 5062 c---aggcctacccttgctcaggttcaagaagacagctgaagagcgctcattccgcattctt 5118
Db 1498 CAGAAGAGATGGCATGGATCAAGAAATCCCCCTTGTGAAGAGATCAACAGCATGTTCTTG 1557
Qy 5119 cttctttgcagaattctcattatatacaagtggtccattgggtgtgatatctcgtgct 5178
Db 1558 CTTTGTATAAACACTTCCACAGTATACGGGAAGTGCTTGAAGAAATACCATCCGGCT 1617
Qy 5179 ggtatttggtggtcactgaagaatggaatcacccttattctcgtgtgctgctcattgt 5238
Db 1618 GGCATCTGTGGAGAGGTCTCAGAGCTGGAATCATGCTTCTCCTCGAAGTGTCTCTCTTAT 1677
Qy 5239 ttttccactatttacttgggtgaactccgcctgaggaactgcataaccaattctgcagaag 5298
Db 1678 TTTTTCATTAATAATGGAGTCTCTCCACCCGACATTCAGT---TCTTGGACAA 1734
Qy 5299 gagagtcagtcgactctgactctatctattttacatacaaatgttctcgtccttcc 5358
Db 1735 GCCATTTTGAACATTTATGTAGCTATCTTCCCTACCAACAACCTCATTTGCCCTTTTC 1794
Qy 5359 aggaatttggtgatactgttaagcccttgctccagaggtggtgctgagatcctccttac 5418
Db 1795 AAGAAATAGTAGATAATGAATTAATCACTGATTTGAAGTGTGTCGCGTAACAGTGAAGTTA 1854
Qy 5419 taaactttgaagcaaaaaaacacccgtggtcaggtaccctagaaaaagaatagtttga 5478
Db 1855 AAGATATCTAGAGGTGAAGAGATGCTATAGATATATCAAGAGATCTAACAAATAA 1914
Qy 5479 tagagctctctgatgactatagctgctcctgaaatcaatcctcatttcaggtgcccac 5538
Db 1915 TAAACCTTCAGAGGATTAACAGAGCTCATTAATCAAGCATCCAATTTCTCGTGCCTGA 1974
Qy 5539 ggtctgcagatgatgagcgaagacatcctgtcctcctctcttctgtggggtatactat 5598
Db 1975 AATCAGGTGCTGATAGAGCAGAGAGCCCAACTCTGTGCTGTGTCGCGATCTCTGCTGT 2034
Qy 5599 gttctgaacatttctgctcaggaataattgtgaacggggaagaggttgagcttgcaattt 5658
Db 2035 GCTCCAGAGTTACTGCTGCACACTGAATGGAAGGGGAGGATGTAGAGCCCTGCACAG 2094
Qy 5659 ttcaagcaacttcaactgtggagcggagctgtcatttctccataaaatcagagaatgccag 5718
Db 2095 CTCACACTACTCTGTGGCTCTGGAGTGGCATCTTCTCTCAGAGTACGGGAATGTGAGG 2154
Qy 5719 tggctcgtggtgaaggtaaagccagagggctgtgctcctatccagctccttacttgatgaat 5778
Db 2155 TGCTATTTTGTAGTGGCAAAACCAAGGCTGTTTTTATTTCTCTCTTACCTTACCTTGATGACT 2214
Qy 5779 atgagaaacagaccctggtcctgaagaggggcaaccccttcttattctcgtgagcgg 5838
Db 2215 ATGGGAGACCCGACAGGAGTCTCAGACGGGAAATCTTTACATTTATGCAAGAGCGAT 2274
Qy 5839 atcggagctcccttggctgtgcgaacaacactgcattatagaagagatgctgtaggagcc 5898
Db 2275 TCANGAGATTCAGAGCTCTGGCACCACACAGTGTTCACAGAGAAATGGACATGCAC 2334
Qy 5899 aagagactaatcagatgttatttgattcaactggca 5935
Db 2335 AGGAAGCAATCAGACTGTTGGCATTCGACTGGCA 2371

RESULT 13
AC068724/c
LOCUS
DEFINITION Homo sapiens' chromosome 15 clone RP11-473C18 map 15q15, complete
sequence.
ACCESSION AC068724
VERSION AC068724.7
KEYWORDS GI:12739735
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 190727)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
Nesbitt,R., Traicoff,R. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
unpublished
2 (bases 1 to 190727)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A.,
Madan,A., Nesbitt,R., Shaffer,T. and Hood,L.
Direct Submission
Submitted (08-MAY-2000) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
3 (bases 1 to 190727)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
Nesbitt,R., Traicoff,R. and Hood,L.
Direct Submission
Submitted (10-FEB-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT
On Feb 10, 2001 this sequence version replaced gi:11138174.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping BACs AC009825 [Drafting center: WIBR]
and AC009852 [Drafting center: UWMSC] were added for finishing
----- Location/Qualifiers
1..190727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-473C18"
/clone_lib="RPCI human BAC library 11"
/note="This clone overlaps RP11-2016 AC021899, RP11-19D14
AC009825 and RP11-402F9 AC009852. Data from AC009825 and
AC009852 were added and the consensus sequence determined
from RP11-473C18 to the extent possible"
1..9704
/note="overlap with RP11-2016, AC021899"
42130..42140
/note="low quality data"
43640..43830
/note="low quality data"
69905..70000
/note="single read coverage"
87320..190727
/note="overlap with RP11-19D14, AC009825"
99800..99900
/note="low quality data"
103370..103380
/note="low quality data"
105748..190727
/note="overlap with RP11-402F9, AC009852"
BASE COUNT 54575 a 40945 c 40610 g 54597 t
ORIGIN
```

Query Match

5.6% Score 354; DB 9; Length 190727;

Best Local Similarity 100.0%; Pred. No. 5 6e-78;

Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 atgcaacaggtccagcaccacacacagatatacaacattcagctaccagagctacgc 482

Db 45831 ATGCAACAGTTCCTCCAGCAGCACCACACAGATATACAAACATTCAGCTACCAAGAGCTACGC 45772

QY 483 ctgataaattagaggggaaaaaaattctccagtccttcacgtcgtagccttgcttc 542

Db 45771 CTGATAAATTAGAGGGGAAAAAAATCTCCAGTCCCTTCAAGTCGTGACGCTTGCTTC 45712

QY 543 cggaaagcggcggcgaagccactctcagctcgtcaaacccacttcaggagccgc 602

Db 45711 CCGGAAGCGGGCCGGAAGCCACTCTCTCAGTCTGCTCAACCCGACACTTCAGGGCCGTC 45652

QY 603 gtaaaagtgtcgtcctctctctcgcacggccacaggtttccctgcttgccttgccctg 662

Db 45651 GTAAAAGTGTCTGCTCTCTCTCGACCGGCCACAGGTTTCCTGCTGTGCGCGG 45592

QY 663 ggtcgcaactgcaggtcgtcagttccctcagatggtcgagagaggtgaggtact 722

Db 45591 GGTGCGCAACTGCAGCGCTCAGTTTCCCTCAAGATGGGGACGAGGAGGCTGGAGGTACT 45532

QY 723 gagaagatggaatcagcggaggttaccacagaccctcagcgtcgtgcatct 776

Db 45531 GAGAGATGGAATCAGCGGGAGGTACCCAGACCCCTCAGCGTGTGCACT 45478

RESULT 14

AC090992

LOCUS AC090992 61896 bp DNA HTG 22-MAR-2001

DEFINITION Homo sapiens chromosome 15 clone RP11-473C18 map 15, LOW-PASS

SEQUENCE SAMPLING.

AC090992

VERSION AC090992.1 GI:13431042

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 61896)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,

Canarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,

Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,

Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,

Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,

Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,

McClean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,

McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L.,

Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,

Savery,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,

Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-MAR-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smt, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12436

Center clone name: 473_C_18

* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1. 716: contig of 716 bp in length
* 717 816: gap of 100 bp
* 817 1540: contig of 724 bp in length
* 1541 1640: gap of 100 bp
* 1641 2352: contig of 712 bp in length
* 2353 2452: gap of 100 bp
* 2453 3172: contig of 720 bp in length
* 3173 3272: gap of 100 bp
* 3273 3996: contig of 724 bp in length
* 3997 4096: gap of 100 bp
* 4097 4805: contig of 709 bp in length
* 4806 4905: gap of 100 bp
* 4906 5633: contig of 728 bp in length
* 5634 5733: gap of 100 bp
* 5734 6441: contig of 708 bp in length
* 6442 6541: gap of 100 bp
* 6542 7240: contig of 699 bp in length
* 7241 7340: gap of 100 bp
* 7341 8065: contig of 725 bp in length
* 8066 8165: gap of 100 bp
* 8166 8845: contig of 680 bp in length
* 8846 8945: gap of 100 bp
* 8946 9638: contig of 693 bp in length
* 9639 9738: gap of 100 bp
* 9739 10463: contig of 725 bp in length
* 10464 10563: gap of 100 bp
* 10564 11276: contig of 713 bp in length
* 11277 11376: gap of 100 bp
* 11377 12086: contig of 710 bp in length
* 12087 12186: gap of 100 bp
* 12187 12901: contig of 715 bp in length
* 12902 13001: gap of 100 bp
* 13002 13695: contig of 694 bp in length
* 13696 13795: gap of 100 bp
* 13796 14509: contig of 714 bp in length
* 14510 14609: gap of 100 bp
* 14610 15328: contig of 719 bp in length
* 15329 15428: gap of 100 bp
* 15429 16140: contig of 712 bp in length
* 16141 16240: gap of 100 bp
* 16241 16955: contig of 715 bp in length
* 16956 17055: gap of 100 bp
* 17056 17773: contig of 718 bp in length
* 17774 17873: gap of 100 bp
* 17874 18561: contig of 688 bp in length
* 18562 18661: gap of 100 bp
* 18662 19396: contig of 735 bp in length
* 19397 19496: gap of 100 bp
* 19497 20214: contig of 718 bp in length
* 20215 20314: gap of 100 bp
* 20315 21031: contig of 717 bp in length
* 21032 21131: gap of 100 bp
* 21132 21845: contig of 714 bp in length
* 21846 21945: gap of 100 bp

/db_xref="taxon:10090"
/cell_type="embryonal stem"
/dev_stage="embryo"
/chromosome="2"
143..399
/gene="Ubr1"
/number=2

BASE COUNT 273 a 150 c 168 g 319 t
ORIGIN

Query Match 3.4%; Score 215.2; DB 10; Length 910;
Best Local Similarity 89.2%; Pred. No. 3.6e-43;
Matches 232; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 777 tgggtggatcagcaagttgattttatatactgcttttgcacatcattggcacaattggtg 836
|||||
Db 143 TGGTGGGATCAGCAAGTTCATTCTATATACTGCTTTCTTACATCATTGGCACAAATTAGTG 202
|||||

QY 837 ccagaaatttactttgctgaaatggaccagacttggaaaagcagagagaaagtgtacaa 896
|||||
Db 203 CCAGAAATTTATTTGCTGAGATGGACCCAGATTTGGAAGCAGAGAGAGAGTGTACAG 262
|||||

QY 897 atgtcaatattcactccactgaatggtagtacttatttggagaagatccagatatattgctta 956
|||||
Db 263 ATGTCAATACTCACCTTTGGAGTGGTACTTATTGGAGAGGATCCGGATATTGCTTA 322
|||||

QY 957 gagaaattgaagcacagtgagcatttcagctttgtggagggttttcaaaagtggagag 1016
|||||
Db 323 GAGAAATTAACACACAGTGGAGCGTTCCAGTTCGTGGGAAGGTTTCAAAAGTGGAGAA 382
|||||

QY 1017 acaacctattcttcagggga 1036
|||||
Db 383 ACAACATATTCTGTAGGTA 402
|||||

Search completed: March 1, 2002, 16:16:46
Job time: 13031 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 10:00:18 ; Search time 49.84 Seconds
(without alignments)
5133.032 Million cell updates/sec

Title: US-09-724-126A-2
Perfect score: 9224
Sequence: 1 MADEAGGTRMEISAEPLQ.....EIARSQETNQLFGFNQQL 1749

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8617	93.4	1757	11 070481	O70481 mus musculus
2	3160	34.3	1275	4 015057	O15057 homo sapien
3	3151	34.2	1273	4 09NU68	Q9nu68 homo sapien
4	2278.5	24.7	1829	5 09VX91	Q9vx91 drosophila
5	1945.5	21.1	1927	5 P91133	P91133 caenorhabdi
6	1716	18.6	333	4 060708	O60708 homo sapien
7	1692	18.3	329	4 075492	O75492 homo sapien
8	606.5	6.6	1225	10 09LZ95	O9LZ95 arabidopsis
9	598	6.5	212	4 09H578	Q9h578 homo sapien
10	549.5	6.0	1872	3 007963	Q07963 saccharomyc
11	537.5	5.8	2168	5 019330	Q19330 caenorhabdi
12	420	4.6	1456	5 09W3M3	Q9w3m3 drosophila
13	343	3.7	795	10 09LZ94	O9LZ94 arabidopsis
14	335.5	3.6	794	10 096248	Q96248 arabidopsis
15	311	3.4	783	5 09W3M5	Q9w3m5 drosophila
16	205.5	2.2	3187	11 063714	Q63714 rattus norv
17	190.5	2.1	3259	4 014789	Q14789 homo sapien
18	188.5	2.0	2041	4 09HAT7	Q9hat7 homo sapien
19	188.5	2.0	2047	4 09UH61	Q9uh61 homo sapien

20	186	2.0	1410	4	Q14221	Q14221 homo sapien
21	185	2.0	1449	6	Q9BG87	Q9bg87 bos taurus
22	184	2.0	1411	4	Q15075	Q15075 homo sapien
23	184	2.0	1979	4	Q95949	Q95949 homo sapien
24	182.5	2.0	2442	4	Q9H450	Q9h450 homo sapien
25	182	2.0	2442	4	Q14812	Q14812 homo sapien
26	181.5	2.0	2442	4	Q60588	Q60588 homo sapien
27	181.5	2.0	2579	11	Q9ESD3	Q9esd3 mus musculu
28	181.5	2.0	2792	11	Q9ESD4	Q9esd4 mus musculu
29	181.5	2.0	2856	11	Q9ESE1	Q9esl mus musculu
30	180.5	2.0	1413	5	Q9XWR0	Q9xwr0 caenorhabdi
31	180	2.0	2244	5	Q9NCG0	Q9ncg0 drosophila
32	179.5	1.9	2346	5	Q01385	Q01385 drosophila
33	177	1.9	1151	4	Q9HCL1	Q9hcl1 homo sapien
34	177	1.9	1416	4	Q9BZF9	Q9bzf9 homo sapien
35	176	1.9	1978	4	Q15154	Q15154 homo sapien
36	176	1.9	2013	5	Q9VKI0	Q9vki0 drosophila
37	174.5	1.9	1009	2	Q99UD0	Q99ud0 staphylococ
38	172.5	1.9	1780	4	Q9UFR5	Q9ufr5 homo sapien
39	171	1.9	994	4	Q9Y489	Q9y489 homo sapien
40	171	1.9	3616	13	Q9W6V0	Q9w6v0 gallus gall
41	168	1.8	1837	3	Q74424	Q74424 schizosacch
42	167	1.8	1578	10	Q9AY25	Q9av25 oryza sativ
43	166	1.8	1676	10	Q23332	Q23332 arabidopsis
44	165.5	1.8	1931	5	Q9NCF9	Q9ncf9 drosophila
45	164.5	1.8	1181	10	Q04650	Q04650 arabidopsis

ALIGNMENTS

RESULT 1

O70481

ID O70481 PRELIMINARY; PRT: 1757 AA.

AC O70481;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE UBIQUITIN-PROTEIN LIGASE E3 COMPONENT N-RECOGNIN (UBIQUITIN-PROTEIN

DE LIGASE E3-ALPHA).

GN UBRI.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,

RA Sangan P., Copeland N.G., Jenkins N.A., Varshavsky A.;

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-446 FROM N.A.

RC STRAIN=129/SVJ;

RA Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,

RA Sangan P., Copeland N.G., Jenkins N.A., Varshavsky A.;

RT "The mouse and human genes encoding the recognition component of the

N-end rule pathway";

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF061555; AAC40165.1; -

DR EMBL; AF067379; AAC23678.1; JOINED.

DR EMBL; AF067371; AAC23678.1; JOINED.

DR EMBL; AF067372; AAC23678.1; JOINED.

DR EMBL; AF067373; AAC23678.1; JOINED.

DR EMBL; AF067374; AAC23678.1; JOINED.

DR EMBL; AF067375; AAC23678.1; JOINED.

DR EMBL; AF067376; AAC23678.1; JOINED.

DR EMBL; AF067377; AAC23678.1; JOINED.

DR EMBL; AF067378; AAC23678.1; JOINED.

DR MGD; MGI:1277977; Ubr1.

DR InterPro; IPR000307; Ribosomal_S16.

DR Pfam; PF02207; zf-UBRI.1.

DR SMART; SM00396; ZnF-UBRI; 1.

RX MEDLINE-97349984; PubMed-9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002347; BAA20806.1; -;
DR InterPro; IPR000408; RCC1.
DR InterPro; IPR001841; Znf_ring.
DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
DR SMART; SM00184; RING; 1.
FT NON_TER 1
SQ SEQUENCE 1275 AA; 145331 MW; 101FF1F66E056066 CRC64;

Query Match 34.3%; Score 3160; DB 4; Length 1275;
Best Local Similarity 47.3%; Pred. No. 5.3e-212;
Matches 612; Conservative 248; Mismatches 393; Indels 40; Gaps 19;

QY 479 LKYLISKPTIWTERTLRMQLEGRSFLKILTCMOGMEEIRROVQGHTEVDPDWEAAIAI 538
DB 1 LKYLISKPTWSEDELKQKLEGEFALELLKCMQMDPIRQVQGHIEPEWEAAFTL 60

QY 539 QMQLKNILLMFQEWCADEELLVAYKECHKAVMRCSTSFSSSKTVVQS-CGHSLKTS 597
DB 61 QMQLKTHVTSMQDWCASDEKVLIEAYKKLAVLMQCHGGYTDGEQPTLSICGHSVETR 120

QY 598 YRVSEDLVSHLPSRLTAGLHVRLSRLGAVSRLEHFVSFDFQVEVLVEPLRCLVLVA 657
DB 121 YCVSQEKVSHLPSVRLAGLHVLLSRSEVAYKPELPLSELSPMPMLIEHPLRCLVLCA 180

QY 658 QVVAEMRRNGLSLISOVYVQVCKREEMVDKDIIMLOIGASLMDPNKFLLLVLORVEL 717
DB 181 QVHAGMRNGFSLVNOIYYHHVVKCRREDFDKDVMLOQGVSMMDPNHFLMILSRFEL 240

QY 718 AEAFTN-----KTSTK--DODLIKQYNTLIEMLQVLIYIVGERYVPGVGNVTKEEVFMR 770
DB 241 YQIFSTPDYKRFSESTHDKDVQONNTLIEMLYLIIMLVGERFSPGVGQVGNATDEIKR 300

QY 771 EIIHLICIEPMPHSAIAKNPENENNETGLNVINKVATPKPGVSGHGVYELKDESLKD 830
DB 301 EIIHQLSIKPMASSELVKSIPDENKETGMESVIEAFAHFKKPLGTGRMYELKPECAKE 360

QY 831 FNMVYFYHSTQSKACHMOKRRKQENKDEALPPPPPPPECPAFKVINLLNCIDIMYI 890
DB 361 FNLIFYHESRAEQSKAEARGLKQRNEDTALPPVLPPECPFLASLIVLQSDVLMCI 420

QY 891 LRTVFERAIDTDSNLMTGMLQAFHIALGLLEEKQLOKAPEEE-VTFDFYHKASRLG 949
DB 421 MGTILQAVEHNGYAWSESMLQVRLHLIGMALQOEKQHLENVTEHVVTFTTQKISKPG 480

QY 950 SSANM---IQMLLEKLGIFOLEGQKDMITWILQMFDTVRLRKS-CLIVATSGSESI 1005
DB 481 EAPKNPSILAMLETQNAPLYLVHKDMIRWLKTFNNAVKMKRESSPTSPVAETEGTIM- 539

QY 1006 KNDIEITHDKAEKPKRAEAARLHRQKIMQMSALQKNEFETHKLMYDNTSEMPCKEDSI 1065
DB 540 --ESSRDKDKAEKRAEATARLRREKIMQMSQMRHFIDENKELFOQTLELDASTAV 597

QY 1066 MEEESTPAVSDYRIALGPKRGSPSVTEKEVLTCLTCQEEQEVKTIENNAMVLISACVQKSTA 1125
DB 598 LDH--SPVASDMTLTALGPTQTQVPEORQVFTCILCQEEQEVKVESRAMVLAFAVQRSTV 655

QY 1126 LTHRGKPIELSGEALDFPMDPLAYGTGTGSGHVMHVAWCQKYFEAVOLSSQ----- 1181
DB 656 LSKNRSFIQ-DPEKYDPLFMPDPLSCGTHTSCTGTHMAHQMRYFDSVQAEQRQR 714

QY 1182 -RIHVDLFDLESCEYLCPCKSLCNTVIPPIPOPKINSINADALAQILLARWIQTVL 1240
DB 715 LRLHTS-YDVNGEFLCPCECNSLNTVIPL-LPPRNFNRLN-FSDQPNLTQWRTIS 771

QY 1241 ARISGYNIRHAKGENPIPIFFNQMGMDSTLEFHSILSFGEVSSIKYSINSIKEMVILFATT 1300

DB 772 QQIKALQLOFLRKEESTPNNASTKNSENVDELQQLPEGRPDPRPKIPYSESKEMLTTFGTA 831
QY 1301 IYRIGLGVPPDRDRPVPMLTWTSTCAFTTQAIENLLGDEGKPLFGALQNRHNGLKALMQ 1360
DB 832 TYKVLGVKHPNEEDPRVIMCWGSCAYITQSIERIISDEDEKPLFGPLPDRDDCLSLR 891
QY 1361 FAVAQRITCPQVLIQKHLVRLLSVLPNIKSEDTPLCLSLDFLHVLVGAVLAFPSLYWDD 1420
DB 892 FAAAHWTVASVSVVQGHFKCLFASLVNDSHEELPCILDDMFHLVLGLVLAFALOCQD 951
QY 1421 PVDLOPSSVSSVSNHLYLFLHITMAHMLQILL---TVDGLPLAQVOEDESEAHSSFF 1477
DB 952 -----FSGISLGTGDLHIEHLVMTMAHIIQLTTSCTEENGMD--QENPPCEESEAVLALY 1004

QY 1478 AEISQYTSGISGCDIP-GWYLVVSLKNGITPYLRCALFEHLGYTPPPPELHTNSAEGE 1536
DB 1005 KTLHQYT-GSALKEIPSGWHLMKRSVRAGIMPFLLKCSALFFHYLNGVPPSPDIQV-PGTSH 1062

QY 1537 YSALCSYLSLPTNLFLLFOEYWDVTRPLQRCADPALLNCLKOKNTVVRYPRKRNSLIE 1596
DB 1063 FEHLCSYLSLPLNNLCLFOENSEIMNSLIESWCNRSNVEVKRYLEGERDAIRYPRESNKLIN 1122

QY 1597 LPDDYSCLLNQASHFRCPRSADDERKHPVLCFPGAILCSQNCQEIIVNGEEVGACIFH 1656
DB 1123 LPEDYSSLLNQASHFRCPKSGGDKSRAPTCLVCGSLLCSQSYCCQTELEGEDVGACTAH 1182

QY 1657 ALHCAGAGVCFIKIRCEVRVVEGKARGCAYPAPYLDVEYGETDGLKRGNPLHLSRERYR 1716
DB 1183 TYSCSGVGIFLRVRECVQLFLAGTKGCFYSPYLDYGETDQGLRRGNPLHLCKBRFK 1242

QY 1717 KLHLVWQOHCIEETARSOETNOMLFGFNQOLL 1749
DB 1243 KIQLWQHQSVEETEGHAQENOTLVGIDWOHL 1275

RESULT 3
ID Q9NU68 PRELIMINARY; PRT; 1273 AA.
AC Q9NU68;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE DJ392M17.3 (K1AA0349 PROTEIN) (FRAGMENT).
GN DJ392M17.3
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Collier R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049843; CAB75421.1; -;
DR InterPro; IPR000408; RCC1.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1273 AA; 145060 MW; A43DCE0BF4C72B39 CRC64;

Query Match 34.2%; Score 3151; DB 4; Length 1273;
Best Local Similarity 47.3%; Pred. No. 2.3e-211;
Matches 610; Conservative 248; Mismatches 393; Indels 40; Gaps 19;

QY 481 YLISKPTIWTERTLRMQLEGRSFLKILTCMOGMEEIRROVQGHTEVDPDWEAAIAIQM 540
DB 1 YLISKPTWSEDELKQKLEGEFALELLKCMQMDPIRQVQGHIEPEWEAAFTLQW 60

QY 541 QLKNTLLMFQEWCADEELLVAYKECHKAVMRCSTSFSSSKTVVQS-CGHSLKTSYR 599
DB 61 KLTHVISMMDWCASDEKVLIEAYKKLAVLMQCHGGYTDGEQPTLSICGHSVETR 120

Query Match		24.7%;	Score 2278.5;	DB 5;	Length 1829;
Best Local Similarity		31.6%;	Pred. No. 4e-150;		
Matches 606;		Conservative 305;	Mismatches 714;	Indels 295;	Gaps 55;
Qy	5	EAGTEREISAEALPQTPQRIASWMDQVDYATFLHHLAQLVPEIYFAEMDPDLEK---	61		
Db	24	QAGTLDRSDIEFLKRESPK---YFDYQTSATVRGAHPRSSRP-----PDVSSFLK 72			
Qy	62	---QEEVSQMSIFPLWELFEGDPDICELEKLKHSQ-AFQLCGRVFKSGETTYSCRDCAI 117			
Db	73	CMFKESLAKEEIIDVVVEFMGLDNPSSALEKLRLLEGNTATVCGKVFKNGEPTYSCREGV 132			
Qy	118	DPTCVLCMDFQDSVHKHNRHKMTSGGCGCDTFAWKTGPECVNHHPGRAGTTKEN 177			
Db	133	DPTCVLCVNGCFKRAHFRHKYKMTSGGGCCDGDDEAKWKQOYCELUHLANRKNPLESK 192			
Qy	178	SRCLNEEIVQARKIPSPYIKYVEMTIWEEREKELPELQIREKN-----ERYVCV 229			
Db	193	I---LTDVAVLERVEICGAILAFVCVSL---ETEPNASLOCLDNGVEGGQVDGAQYCTV 245			
Qy	230	LFNDEHSYDHVYISLQALDCELAELAQLHTTADTKGGRVAVKAGAVAAOEAKEDEKSH 289			
Db	246	LYNDESHTFDQVQTLTKIAKRAKADAMEIVAAIDREGRAVVKCDTPEECNKLKVSIEHQ 305			
Qy	290	-----SENVSQPLHVEVLHSEINAHOKFALRLGSMNKMINSYSSDPRQIEQACLR 341			
Db	306	MILPSTSVTARNQSLRTSVLHIGAVACQOAFALQLLGNFQFELVRHYLFKRTFSELVQR 365			
Qy	342	EEPDSENPCLISRLMDAKLYKGARKILHELIPSPFEMEMEYKKLFAMEYKYKQLQK 401			
Db	366	KQ---ETFC-IRHILEYDVKLWKTARTCWHRLLLSGLMEYDNKMLIAQESRYAYIVE 421			
Qy	402	EYISDDHRSISALTASVQMTVPTLARHLIEBQNVTSVITETLLEVLPEYLDNRKNFN 461			
Db	422	DFISDDHDHAFSVLSVQLFTVPSIAHLIAHEGIFDKLLIHYHVAIEKFIKNTLHF 481			
Qy	462	QG---YSQDKLGRVAVICDLKYLISKPTIWTETRLRMQFLEGRSFLKILCTMOMGEIR 519			
Db	482	SKNTASITFKRANVILYDURLYLSLPDVLNDLNGFLGGCRALMRVLNVMQGESMT 541			
Qy	520	RVQGHIEVDPDWEAAIAIOMQLKNILLMFQEWACACDEELLLVAYKECHKAVMRCSTSF 579			
Db	542	RTQGHMDYEPWECAFNHLKLTATTSQVIDNASGDVKLLKLYKMTMRALV--SNSFI 599			
Qy	580	SSKTVVQS---CHSLETYSRVSDSLVSIHLPLSLTLAGLHVRLSLGAVSRHLFEVS 636			
Db	600	VGEKVMQPKKVDHVAIVLVISVQVPSVTHPLSRFFAGIYL---HLGA---HD-LT 651			
Qy	637	FEDFQVEV-----LVEYPLRCLVLVAQVVAEWMRRNGLSLISQVFFYQDVKCRE 685			
Db	652	YDGLQTEALSILKTPREIIEPVLCVQAMTAQVAGLRRNGYTLHLHLYFYFNRVCRV 711			
Qy	686	EMYDKDIIMLQIGASLMDPNKFLLLVQRYE---LAEAFNKITSTKD--QDLIKQYNTL 739			
Db	712	EMLDRDIACIQIGASLMESEFLHVLNRNTIPWLQENTWLSLSGNEMNDIIRE-ASI 770			
Qy	740	IEMLQVLLIVGGRYVPGVGNVTKEETMREIHLHCIEPMPHSAIAKLNLPENE--NNE 797			
Db	771	FDEPLELLIIVIGRMWPGVSMVTEEDRLRKEIIQLLCKIPYSHSELSSRALPDGNSGD 830			
Qy	798	TGLENVINKVATPKK-CVSGHGVYELKDSRLKDFNMFYHYKTSQHSKAEHMOKKRRQ 856			
Db	831	NVFEVINTVAVFKPKVGADSKGVYELKEHLLEKFNMFYHYTKEDSKAEELQERRKA 890			
Qy	857	ENKDEALPPPPPPFCFAFKNLLNCDIMMYILTRVVFERRAIDTSDNLNLTGMLQMAFH 916			
Db	891	KKQLVCCPPMLPKLTFTPTPMANILQCPVFLNCSLIMEALNAYSRSTFESHLQVLH 950			
Qy	917	ILALGLLEEKQOQLOKAPEEVTFDYHKAISRLGSSAMNIQMLLEKLGIPQLEGQKDMIT 976			
Db	951	LLGVAIQEE-----LSEHYFPFLSFYERSQEYG-----ILEKLEELARCPRLEAHYDFVL 999			

Qy	977	WILQMFDTVKRLREK-----SCLIVATTSGSESKNDEITHDKKAEKRKAEEA 1026
Db	1000	WTIERF---KLOAKOAPSDGRAGPSC-SQOQTGGKLSLAE---QAREERENARLA 1051
Qy	1027	RUHQKIMQAQMSALOKNFTHKLMYDNTSEMPGKE-----DSIMEESTPAVSQY 1077
Db	1052	AERRAHIMAQMAKQKSFISNAEMFADENETRKESASTGPMDEWEDIPPEEQGAVAL 1111
Qy	1078	SRIA-LGPKRGFSVTEKEVLTCTILCOEBOEVKTIENNAMVLSACVQKSTALTQHRKPIEL 1136
Db	1112	SKVACLGDPDRKPYHGTDDTFKILCFENCAISRGGQLVSSAFVQTSRV----- 1160
Qy	1137	SCEALDPLFMDPDL---AYGTVTSGCHVHVAWCWKYF---EAVOLSSQORIHVDLFDL 1190
Db	1161	-----TFTTNLRNSQALHISCCGHVHYSWLEYFTNEEFKELRRPHRRAALAAQ 1213
Qy	1191	ESGYLCPCLCKLONTVPIPILOPO-----KINSENADALAOL--- 1229
Db	1214	ANVEFQCPYCRTLNSNAIIPVETLPAFSAPPSNESYPLDPSFEIMSTLAIELGNVKDH 1273
Qy	1230	-LTLARWQTVLARISGYNIRHAKGENPIPIFNQMGDSTLEFHSILSPGVSESSIKYSN 1288
Db	1274	ELTTLPSVSNIL-RLSGVVGGLAQFERSQLIKNPP-----RLHADYIEGIE----- 1319
Qy	1289	SIKENVILLFATTIYRIGLK-VPPDRDRPVPMLTWTSCAPTIOAIENLLGDECKPLFGAL 1347
Db	1320	-FLKALLNTMKIQOHLKADHPAIESIEMVPII-WDSCSYTLOALEIYLVAVKPKLAKEL 1377
Qy	1348	QNRHONGKALMQ-----PAVAQRITCPQVLIQKHLVRLSVLPLNPKISBDTPCLLSIDL 1402
Db	1378	SMRHQSCARNLVRACSRSSALEWETDPLLPMPRSQAESFSSRLDITFNQNDTSVLEWDC 1437
Qy	1403	FHVLV---GAV-LAPPSLYWDDPVDLQPSVSSVSNHLYLFLHLITMAHMLQLLFTVDTG 1457
Db	1438	FRVLVPFQFGLNLMVPEKGYKTIIP-----SGSMFDFYIMOTMFLAQLTKAVLCFDVE 1491
Qy	1458	LPLOA-----VOEDSE 1468
Db	1492	KEKARAEKAPNSELTQDIYIQLPSRIRDNMIDFYRRYINIPARVLQTKQKOLVEESE 1551
Qy	1469	E---AHS-----ASSFFAEISQYTSIGSICDIPGWYLMWSKNGITPYLRCAALFFHYL 1519
Db	1552	ENQHGQTVVIPCESHLLALLEY-----VQOMSSFLRCSCLFYRFL 1594
Qy	1520	LGVTPEELHTNSAGEYSALCSYLSLPTNLFLLFQBYWDTV-RPLLRWCADPALLNCL 1578
Db	1595	TDVDFPDPPTDQPD-RFDLMQCYLGLDPMGLGVYFD--METVYATMMHFSAPHIDREV 1651
Qy	1579	KOKNTVRYPRKNSI-IE-----LPDDYSCLLNQASHFCRPSADDERKHPVL 1626
Db	1652	EOR-----COPDARRSLQVEPCLRPRLKVLCDDFSDLINSVSDIFCPNNREEMKPTTM 1707
Qy	1627	CLFCGAILCSQNICQEIYNGEVEGACIFHALHCGAGVCIIFKIRRCRVVLVVEGKARGCA 1686
Db	1708	CLICGLILCGOSYCCOPELGKYSVGACTHHAACGAEGVIFLRIDCQVYVL-GRGKGF 1766
Qy	1687	YPAPLYGETDTPGLKRNPHLSRERYKHLVWQOHCIIIEIARSQETQMFLGFNW 1746
Db	1767	VPPLYDEYGETDMGLRRGNLRLSQAAVRYQLWHLGHLGHEIARLNDNANVAAAAQW 1826

RESULT 5
P91133
ID P91133
AC P91133; PRELIMINARY; PRT: 1927 AA.
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILAR TOS. CEREBITIAE UBIQUITIN-PROTEIN LIGASE E3 COMPONENT
GN SP:P19812.
GN C32E8.11.
OS Caenorhabditis elegans.

Db 1454 GFEEPTKDLGKNMKFRKRGNELKTNFTFEKHLKGVIVSTVWTQSTAHVARAISSYLHYDK 1511
 Qy 1341 KPLFGALQNRHNGKAL-----MQFAVAQRITCPQVLIQKHVR--LLSWL--- 1386
 Db 1514 KPLFGALNRQDCLSAMARLCASLSHNQFLL-----HAYSDMLRVFLCEP 1560
 Qy 1387 PNTKSEDTFCLLSIDLFIHVLGVAGVLAFFSLVWDDPDVLOPSSV--SSSYNHLYLFHLITMA 1445
 Db 1561 PRPKLAQTP-----GSPLLSAPSTSFTEPA---PAQIPHSGTNFAFLVQLENPA 1606
 Qy 1446 -----HMLQILLTVDGTGLPLA-----OVQEDSEEAHSASSF 1476
 Db 1607 GPRKNYNMLNQIDT-----LSLAITHSEADGNDVNMEQESQMEVDVPAQAOTIRKL 1661
 Qy 1477 FAEISYVTSGSGICDIPGWYLVWSLKGITPYLRCAALFFHYLLGVTPPEEL---HTNSA 1533
 Db 1662 YALCHPFDGLRRVDI---LWRKMEGAQSLLRPAILLYHFIITLVDPPFALKDPSINSS 1717
 Qy 1534 EGEYSALCSYLSLP-----TNLFLFQYDWTVRPLLRQWCADPALLNCLKQK 1581
 Db 1718 E----PLFRYLGLPHKIBEQISGSMLEKFTM---WSSSIP-----SDQAL-----R 1757
 Qy 1582 NTVVYPRKRNLSIELPDDYSCLLNQASHFRCPSPADDERKH---PVLCFLGCGAILCSQNI 1639
 Db 1758 QDVLQVQVRPNLLVPEPKYSOLINQVATFCPTIPIBESTSNVPTLCLCVGTILCSQAY 1817
 Qy 1640 CCOEIVNGEVCACIFHALHCGAGVICFLKTRCVRVLVECKARGCAYPAPYLDEYGETD 1699
 Db 1818 CCKIINKQSYGACRYHMSQSGSVGMFLRVNDCSLVMTTRKRGCFRPPYVDFEGEVD 1877
 Qy 1700 PGLKRGPNLHLSRERYRKLHLVWQOHCIIETARQSQETNQLGFNW 1746
 Db 1878 QGFRRCGNLHLNPELYQKLSLWQQGITEEVVNYNEIDFRNVQYDW 1924

 RESULT 6
 060708 PRELIMINARY; PRT; 333 AA.
 AC 060708;
 DT 01-AUG-1998 (TreeBLrel. 07, Created)
 DT 01-AUG-1998 (TreeBLrel. 07, Last sequence update)
 DT 01-JUN-2000 (TreeBLrel. 14, Last annotation update)
 DE UBIOUITIN-PROTEIN LIGASE E3-ALPHA (FRAGMENT).
 GN UBRI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 ON NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,
 RA Sangen P., Copeland N.G., Jenkins N.A., Varshavsky A.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF061556; AAC39845.1; -
 GN Ligase.
 FT NON_TER
 KW 333 333
 SEQUENCE 333 AA; 38541 MW; C8BBB8B9EEF4ADEE CRC64;

 Query Match 18.6%; Score 1716; DB 4; Length 333;
 Best Local Similarity 99.7%; Pred. No. 6,5e-112;
 Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps

 Qy 702 MDPNFKLLVLQRYELAEAFNKTISTKDODLIKQNTLIEEMQLVLIYIGERYVPQVGN 761
 Db 1 MDPNFKLLVLQRYELAEAFNKTISTKDODLIKQNTLIEEMQLVLIYIGERYVPQVGN 60

 Qy 762 VTKEEYVTHREIHLICIEPMPSAIAKNLPENNETNGLVINKVATFKKPGVSGHGV 821
 Db 61 VTKEEYVTHREIHLICIEPMPSAIAKNLPENNETNGLVINKVATFKKPGVSGHGV 120

 Qy 822 ELKDESLKDFNMVYFHSYKTOHSAEHMQKRRKROENKDEALPPPPPPCFAPFSKVINL 881

Db	121	ELKDES	LKDFNMYYFYHYSKTQHSKAHMQKRKKQENKDDEALPPPPPFCFAFSKVINL	180
Qy	882	LNCDIMY	ILRTVFERAIDTDSNLWTEGMLQAFHLIALGLLEEKQOLKAPPEEVTFDF	941
Db	181	LNCDIMY	ILRTVFERAINDTSNLTWTEGMLQAFHLIALGLLEEKQOLKAPPEEVTFDF	240
Qy	942	YHKASRLGSSAMNIOMLLEKLKIGIPOLEGQCKDMITWLQMFDTVRRLREKSCLIVATTSG	1001	
Db	241	YHKASRLGSSAMNIOMLLEKLKIGIPOLEGQCKDMITWLQMFDTVRRLREKSCLIVATTSG	300	
Qy	1002	SESIKNDEITHDKAEAKERKAAEARLHRQKIM	1034	
Db	301	SESIKNDEITHDKAEAKERKAAEARLHRQKIM	333	
RESULT	7			
ID	075492	PRELIMINARY;	PRT;	329 AA.
AC	075492;			
DT	01-NOV-1998	(TrEMBLrel. 08, Created)		
DT	01-NOV-1998	(TrEMBLrel. 08, Last sequence update)		
DE	01-JUN-2000	(TrEMBLrel. 14, Last annotation update)		
DE	UBIQUITIN-PROTEIN LIGASE E3--ALPHA (FRAGMENT).			
GN	UBR1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Kwon Y.T., Reiss Y., Fried V.A., Herssho A., Yoon J.K., Gonda D.K.,			
RA	Sangan P., Copeland N.G., Jenkins N.A., Varshavsky A.;			
RT	"The mouse and human genes encoding the recognition component of the			
RT	N-end rule pathway."			
RL	Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AF067387; AAC23677.1; JOINED.			
DR	EMBL; AF067380; AAC23677.1; JOINED.			
DR	EMBL; AF067381; AAC23677.1; JOINED.			
DR	EMBL; AF067382; AAC23677.1; JOINED.			
DR	EMBL; AF067383; AAC23677.1; JOINED.			
DR	EMBL; AF067384; AAC23677.1; JOINED.			
DR	EMBL; AF067385; AAC23677.1; JOINED.			
DR	EMBL; AF067386; AAC23677.1; JOINED.			
KW	Ligase.			
FT	NON_TER	329		
SEQ	SEQUENCE	329 AA; 38026 MW; 4656ABADDF099BC CRC64;		
Query Match	18.3%; Score 1692; DB.4; Length 329;			
Best Local Similarity	99.7%; Pred. No. 3e-110;			
Matches	328; Conservative	0; Mismatches	1; Indels	0; Gaps
Qy	702	MDPKFLLLVLYQYELAEAFNKTIISKDODLIQYNLTIEEMLQLWIYIVGRYYPGVGN	761	
Db	1	MDPKFLLLVLYQYELAEAFNKTIISKDODLIQYNLTIEEMLQLWIYIVGRYYPGVGN	60	
Qy	762	VTKEEVTMRIILLICIEPMSATAKLNLENNETGLENVINKVAFKPGVSCHGY	821	
Db	61	VTKEEVTMRIILLICIEPMSATAKLNLENNETGLENVINKVAFKPGVSCHGY	120	
Qy	822	ELKDES	LKDFNMYYFYHYSKTQHSKAHMQKRKKQENKDDEALPPPPPFCFAFSKVINL	881
Db	121	ELKDES	LKDFNMYYFYHYSKTQHSKAHMQKRKKQENKDDEALPPPPPFCFAFSKVINL	180
Qy	882	LNCDIMY	ILRTVFERAIDTDSNLWTEGMLQAFHLIALGLLEEKQOLKAPPEEVTFDF	941
Db	181	LNCDIMY	ILRTVFERAIDTDSNLWTEGMLQAFHLIALGLLEEKQOLKAPPEEVTFDF	240
Qy	942	YHKASRLGSSAMNIOMLLEKLKIGIPOLEGQCKDMITWLQMFDTVRRLREKSCLIVATTSG	1001	
Db	241	YHKASRLGSSAMNIOMLLEKLKIGIPOLEGQCKDMITWLQMFDTVRRLREKSCLIVATTSG	300	
Qy	1002	SESIKNDEITHDKAEAKERKAAEARLHR	1030	

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003442; AAF46302.1; -.
DR Flybase; FBgn0029985; CG1531.
DR InterPro; IPR001841; ZnfRING.
DR SMART; SM00184; RING; 1.
SQ SEQUENCE 1456 AA; 161358 MW; 61C55D04E4067893 CRC64;

Query Match 4.6%; Score 420; DB 5; Length 1456;
Best Local Similarity 21.3%; Pred. No. 2.6e-20;
Matches 210; Conservative 135; Mismatches 342; Indels 298; Gaps 41;
QY 896 ERAIDTSDNLWTEGMLQAFHILALGLLEEKQLOKAPAEVTFDYH-KASRLGSSAMN 954
DB 599 DSTMDVDCN---EASTSMA-----AAESTALAEGRKRNKNIHVSSKRIGDGPFF 647
QY 955 IQMLEKLGIPQLEGQDKMITWILQMFTVK-RL-----REKSLIVATTSSEIKNDE 1009
DB 648 IGNLLRKI-----AKQD--EQCAQSIDDIRALWPNQREKQ---AAKAREA----- 689
QY 1010 ITHDKEAERKKAAPARLHRQIKMAQSAQKNIETHKLMYDNTSEMPGKEDSIMEE 1069
DB 690 ----KEKEERKK---AREQCKMQDFANKLFQWQSAASGSGMGYPEDDEDELYE 742
QY 1070 STPAVDYSRIALGPKRGPSVTEKEVLCILCOEQEVKNIENAMVLSACVOKSTALTOH 1129
DB 743 EQPREKEYD-----CIICNCTTP-STESNPGLVVLVE--SSGIVGH 781
QY 1130 RGK-----PIELSGEALDPL-----FMDPDLAY--GTVT 1156
DB 782 RRRIARPLPINADESLAHTTRLAEFNRRTTELLSLKFGDESWYLSNNWAYDNGVIV 841
QY 1157 GSCGHVMHVCQKQYFEAVQLSSQRIHVLDLEDSEGYCLPCKSLCNTVPIIPL--Q 1214
DB 842 QSCGHVHLSCLEAYLKTLYTQRPVQ-----DRGEYFCVPCVQLSNSVLPQLDR 895
QY 1215 PQKI-----NSENADALQLLTLARIQTVLARISYNIHAKENPIPIFFNOGMGDSTL 1270
DB 896 PTHLVRSNGQPPERVLADTLDIKENETI-----PQPTKLTAMGHAM- 938
QY 1271 EFHSILSGVESSIKYSIKEMVILFAITVIYRIGLKV-----PDERD 1314
DB 939 ---EVMYIAQRKVCSSITFRKLFIVTSIARTNLEAEIIRGGSLCTANATRYKPKRE 995
QY 1315 PRVPM-----TWSTC-----AFTQIAENLLGDEGKPLFG 1345
DB 996 CIVPLLVHLSVHVRLVVEWPLWSSNASLAGLPVTATEPLPAHCLLELISILADP----- 1049
QY 1346 ALQNRQHNLKALMOFAVAQRITCPOVLQKHLVRLLSVVLNPKSEDPCLLSIDLFIH 1405

DB 1050 -----TALLKFI-----LLAPLQLDQDYFTCMKV-----MYNLLYQI 1084
QY 1406 LVGAVLAPPSLYWDPDLQPSVSSSYNHLYLFHLITMAHMLQ---ILLTVDTGLPLAQ 1462
DB 1085 VVOLCVTLTDLECDHIVKVGSTVSGSDN-----SAAESQQQESAAGTTNNRRRAQ 1136
QY 1463 VQEDSEAHASASSFFAISO-----YTSIGICDIPGWYLVWSLKNKIT--- 1506
DB 1137 QOQSSSLQGLKAMALV--LSQTNDLVHLRRDCIPSTSSAAAAGSSSTTSTNHGASAAT 1194
QY 1507 -----PYLRCAALFFHYLLGVTPPEELHTNSAEG-EYSALCS 1542
DB 1195 ASSATTIEVNLKSMELQALCLFLRVAALLRQHLHYHEMPE-----ISAPGLEFVRLVY 1250
QY 1543 YLSLPTNLLFLFOEYWD-----TVRPLLQRWC-----ADPALLNCLKQ 1580
DB 1251 YLELVTD-----SMDWDCFNASKGLCFIPGTETTLQPWCQQLMEVRPPADTVRELVLIN 1305
QY 1581 KNTVVRYPKRNLSLIELPDDYSCLLNQASHFRCPRSADDERKHPVLCFCGAILCSQNIC 1640
DB 1306 QHSLWQQPR---LLELPREYERLFTY-YHERPCLNCYKVPKKESSICLLCGTIVCLKQNC 1360
QY 1641 COELVNGEEVGACIFHALHCGAGVCIFLKIRECRVVLVEGKARGCAYPAPYLDEYGETDP 1700
DB 1361 CAE-----NDCCEAVRHTLSGCGGIGIFLVTTSTIIIVIRGR-RACLWGLYLDDEEDR 1415
QY 1701 GLKRGNPLHLRSRERYRKLHLWQOH 1725
DB 1416 DLKRGKPLYLKSKDRFNLESQWLSH 1440
RESULT 13
Q9LZ94 PRELIMINARY; PRT; 795 AA.
ID O9LZ94;
AC O9LZ94;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE ECERIFERUM3 (CER3).
GN T1E22.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A., Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALI62874; CAB85536.1; -.
DR InterPro; IPR002114; PTS_HPR_ser.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
SQ SEQUENCE 795 AA; 89971 MW; 401B2D08514B847D CRC64;
Query Match 3.7%; Score 343; DB 10; Length 795;
Best Local Similarity 23.0%; Pred. No. 2.5e-15;
Matches 190; Conservative 123; Mismatches 274; Indels 240; Gaps 43;
QY 1033 IMAQMS-----ALQKNFTETHKLMY-----DNTSEMPGKEDSIMEESTPAVDYSRIA 1081
DB 56 LMLQSLDDTTSESANMIESIKARLIGNQTEKRSDDGRGKDESNKE---SLEIAMYQIVR 112
QY 1082 LGPKRGPSVTEKEVLCILCOEQEVKNIENAMVLSACVOKSTALTOHGRKPIELSG--- 1138
DB 113 -----NKTENINQSLTRVDHQPHEAEN-----CSEKNSV-----GGPSTLQGRFP 153

DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DE CG1530 PROTEIN.
GN CG1530.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephyroidae; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003442; AAF46300.1; -
DR FlyBase: FBgn0029983; CG1530.
DR InterPro: IPR003126; Znf-UBR1.
DR Pfam: PF02207; zf-UBR1; 1.
DR SMART: SM00396; Znf-UBR1; 1.
SQ SEQUENCE 783 AA: 88126 MW; ED0F6FDDFA85C6E4 CRC64;

Query Match 3.4%; Score 311; DB 5; Length 783;
Best Local Similarity 20.5%; Pred. No. 4.3e-13;
Matches 137; Conservative 122; Mismatches 241; Indels 168; Gaps 28;
QY 57 PDLKQESVQMSITP-----LEWYLF-----GEPDICLEKXHSAGFQLGRVF 103
DB 173 PDL----DSIMDTLPNPGTFIDNLONIEWRLIAGGRTPQEFVKIVSYDNHAKCGLVW 228
QY 104 KSGETTICRDAIDPTCVLCMDCFDSVKNHRYKWHSTGTGGCGDCGDTAMKTPGFC 163
DB 229 VPHVYVAYRCITCGISPCMSICRDCFKKGNHTNDFNFLSQAGGACDCGDTSVNKAEGFC 288
QY 164 VNHEPGRAGTICKNSRCPLNEEVIVQARKIFPISVIKYVEMTIWEEKELPPELQIREKN 223

Db 289 SDH-----GINNVNRDPVNNLLIATAVAEAIMPKLLFRLLQH--FREHSTPLEVOAITS- 340
QY 224 ERYCYVLFNDEHSHSYDIYVYISQIRALDCELAQAQLHTTAIDKRRRAVRKAGAYAAQCEAK 283
Db 341 --YSCFEFANMLIDLNNNGEIMRKVMVTRTLINPEYVAFME-----APQDTR 386
QY 284 ED--IKSHSENVSOHPLHVEVLHSEIMAHQKFPALRLGSMNMKIMSYSSDFRQFCQACLR 341
Db 387 NGREFKANRE-----KYEDAVNRFPNP----- 408
QY 342 EEPDS--ENPCLISRLMLWDKLYKGARKILLHELIFSFEME-----MEY 384
Db 409 EPPDEYRDLPALGDKLV-----HTLLGEFTFTFKFEFPQTLVCFLLNMLPDQDY 459
QY 385 KKLFAFMEVKKYKOLQK--EYISDDHDSISITALSVMQFTVPTLARHLIEQNVISVIT 442
Db 460 KEHLTRTFVMHYSRIPSVLEMSRDPDTLSNRVHMSVQLFSNESLAKMWNLSLLHVM 519
QY 443 ETLLVLPPEYL-----DRNNKFN-----QGYSDKLGRRVYAVICDLKYILISKPTIWT 491
Db 520 ISLKLMMSKILIQNTLTHDPNKNFHVICTRQVMKDHG--YWPLVSDFNVLH----- 571
QY 492 ERLRMOFLEG---PRSFLLKILTCMOGMEETRRQVCOHIEVDPD--WEAAIAIQM-----L 542
Db 572 ESVALVFLRDLNLDIMWFQFLOLQGMNVNVRRETASHVEFEPNSYAAFSCELEASAYPM 631
QY 543 KNILLMFOE-----WC--ACDEELLVAYKECHKAVMRCSTSFISSTKTVQS 588
Db 632 WSIISHLODGTGHAHLAKKIINYCVTTLHWDLSIYFMEARLSWVGIFNPIAVS----- 685
QY 589 CGHSLETYSY--RVSEDLV--SIHLPLSRTLGLHVR--LSRLGAVSRLEHFVSFEFQVE 643
Db 686 -----IKSFYLSVQOEMMQASFHPLHRYLAAAFVCQAVTKMGI--SLNDVLPSPRPYLLP 737
QY 644 VLVEYPLR 651
Db 738 LLMIHPLR 745

Search completed: February 28, 2002, 10:05:27
Job time: 309 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	1038	11.3	1958	1	UBRL_SCHPO	O60152	schizosacch
2	914	9.9	2052	1	YDOB_SCHPO	O13731	schizosacch
3	689.5	7.5	1950	1	UBRL_YEAST	P19812	saccharomyc
4	647.5	7.0	1941	1	UBRL_KLULA	O60014	kluyveromyc
5	208	2.3	2230	1	GOG4_HUMAN	Q23439	homo sapien
6	180.5	2.0	1829	1	MY5A_CHICK	Q02440	gallus gall
7	175.5	1.9	1427	1	REST_HUMAN	P30622	homo sapien
8	175.5	1.9	1526	1	MY52_SCHPO	Q9us16	schizosacch
9	175.5	1.9	2663	1	GENE_HUMAN	Q02224	homo sapien
10	170	1.8	2298	1	C215_HUMAN	Q9Y375	homo sapien
11	168.5	1.8	3911	1	AK98_HUMAN	Q95996	h a kinase
12	164	1.8	2096	1	BP28_DROME	Q9vm75	drosophila
13	163	1.8	1853	1	MY5A_MOUSE	Q99104	mus musculus
14	158.5	1.7	1855	1	MY5A_HUMAN	Q9Y411	homo sapien
15	158	1.7	2033	1	EVPL_HUMAN	Q92817	homo sapien
16	156.5	1.7	1790	1	USO1_YEAST	P25386	saccharomyc
17	156.5	1.7	2710	1	TOXA_CLODI	P16154	clostridium
18	156.5	1.7	2869	1	RBPL_PLAVB	Q00798	plasmodium
19	155.5	1.7	1216	1	P1B1_RAT	P10687	rattus norv
20	155.5	1.7	2704	1	BPA1_HUMAN	Q03001	homo sapien
21	155	1.7	3685	1	DMD_HUMAN	P11532	homo sapien
22	153.5	1.7	1875	1	MLP1_YEAST	Q02455	saccharomyc
23	153	1.7	1682	1	MSPL_PLAF3	P19598	plasmodium
24	152.5	1.7	1325	1	G160_MOUSE	P55937	mus musculus
25	152.5	1.7	3433	1	UTRO_HUMAN	P46939	homo sapien
26	151	1.6	3660	1	DMD_CHICK	P11533	gallus gall
27	151	1.6	5430	1	ACF7_HUMAN	Q9unp3	homo sapien
28	150.5	1.6	1539	1	Y373_HUMAN	O15078	homo sapien
29	150.5	1.6	2469	1	TEGU_HSVSA	Q01056	herpesvirus
30	150.5	1.6	5327	1	ACF7_MOUSE	Q9qz20	mus musculus
31	149.5	1.6	1453	1	Y373_BOVIN	Q9tu23	bos taurus
32	149.5	1.6	1679	1	Y109_YEAST	P40457	saccharomyc
33	149.5	1.6	2017	1	MY5N_DROME	Q99323	drosophila

QY 100 GRVKSGETTYSRCDADPTCVLCMDCFDQSVHKNHRYKMHTSTG-GFCDGCDTEAWK 158
Db 97 GHIFRKGVEFYRCCTSVDSNSALCVKFRATSHKHDTSTFVSAGSGCCDCGNAAWI 156
QY 159 TCFPCVNEPGRAGTIKENSRCPLNEEVIQARKIFPVSIVKVVEMTWEER--KELPPE 216
Db 157 GDVSKTHSHEEDATISNDMIDEIPEKLENSIQTTIDCVLDFVLDVFCSPENLKKMPTL 216
QY 217 LOI--REKNER-----YCVLFNDEHHYSVHVIYSQRLACELAE--A 256
Db 217 ESILQDEKTSRLSENKYGDIDDSNMYSLVLWDEKHSFKQFQEQITTAI--ELPNVFG 274
QY 257 QLTHTAIDKEGRRAKAYACQAEKEDIK---SHSENVSOHPLHVEVLHSEIMAHQKF 313
Db 275 KKMAMIINDIGR-----ACIVTETNIKELKIGOKLAQINLAVSIRSMRDFRES 325
QY 314 ALRGLSMNKMITSYS-----SDFRQIFQOACLR----- 341
Db 326 CAVLEWLADIAGSICGRNYFSSVICKELVRPWCNGLHNSDLTFRLSRLALPEIVA 385
QY 342 -EEDP-----SENCLISRLM----- 356
Db 386 IDSPDIFLNEDHINSGPSDSTSHMLETDESSIHSHRWYPSNSLPDVLASYASRVRFDPF 445
QY 357 LWDAKLYKGARKILHELIFSFFMEMEKKLFAFMEFYKYLQLOKEYISDDHRSISITA 416
Db 446 LYDLKMLSLRYKLOELVGYFITQPGKEIMGARIAISYRLAELFLLDREPHSVIF 505
QY 417 LSVQMTPTLARLHIEBQNVISVITETLLEV-----PEYLD-----RNNKENFGY 464
Db 506 FSMQIFTVADVAKLLVTEYDFLTINATLYTFYFKLNTPNYVDQHAMI RTDSAAPH-- 563
QY 465 SODKLGRVYAVICDLKYILISKPTIWTBLR--MOFLEGFRSFLKILTCMOGMEIRQV 522
Db 564 -----SRRYIHIFHHIQMPLSPCV-AEIVREDLKFLQYADFFNL---FGCMCPYTRAV 614
QY 523 QOHEVDPD-WEAATAIQWL-----KNLLMPQEW-----CACDEELLVYAYKECHRAV 571
Db 615 SCHWEENDSMYVNLVSLQVAKCRHGVNFMELNTKLANAINYLISLILYPKARNES 674
QY 572 MRCSTSFSTSSKTVVOSGCHLSKTSYRVEDLSVSIHLPLSRTLGLHVLRLGAVSRL 631
Db 675 W-TWESLTGTTVDGR--GNS-KLIEYDIALQPVSFHPLHLVLV-----L 719
QY 632 HEFVSFEDF-----QVEYL--VEYPLRCLVLVAQVVAEMRRNGLSLSQVFFYQDVCR 684
Db 720 SFYVERONYKLLTQDLAVTDHPLRVCAWLSQMRKALWIRNGTTLRDAQAHYRNLSFH 779
QY 685 EEMYDKIIMLOIGASLMDPNKFLLLVQRYELAEAF--NKTISTKODDLKQYNTLIEE 742
Db 780 EYTFDLVLLQLTLTYGDPDAILPFSIRFQLEDQMGYRFFVPHKHVD-VSQVTIMMEE 838
QY 743 MLOVLIYIGERYPGVGNVTKEEVTMETEITHLICIEPMPSAIAKNLPENNETGLEN 802
Db 839 FLLLLISIVCNVANTLDHWDITRR--IEVGIHILCFRLPYPSEITKRTCEHLEHKKQES 896
QY 803 VINKVATEKK--PGVSGHGVYELKDESLDFNNFYHYSKTSQHSKAEHMKKR--RKQENKD 860
Db 897 TLKQVATERNAEGINDSGSFTLKDEYFDYVDPFNHLYSRNQREAEANILRRYSKQSHKH 956
QY 861 EALPPPPPPPCPAFSK---VINLNCIDIMMYIL-RTVFERAIDTDSNLWTEGMLQMAFHI 917
Db 957 --LESVVYEEYHPIHLSNITIPILQSDSFVGLWHTIYVAYIYPYDQGLKGLVNTALHA 1014
QY 918 LALGLEEKQOLQKAPEEVTFDYHFKASRLGSSAMNTOMLEKLGIPQLEGOKD---- 973
Db 1015 CLLVLMSEK-----GSEPIFSKICE--NRFPVVEGLEQYCN 1050
QY 974 ---MITWILQW-----FDTVKRLREKSLIV-----ATTSGSEIKNDE 1009
Db 1051 PDVTLFVSIVCMKKHNRNFVYVK---EKISLIMKILKSEVPLLYEPVYAEIUSISSKIVQ 1107
QY 1010 ITHOKEAERK--RKAEARLHROKIMAQMSALQKNFIETHKLMYDNTSEMPGKEDSIMEE 1068

RESULT 2

YDOB_SCHPO
ID YDOB_SCHPO STANDARD: PRT: 2052 AA.
AC 013731;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 234.0 KDA PROTEIN C15A10.11 IN CHROMOSOME 1.
GN SPAC15A10.11

Qy	504	SFKILUTCWQMEETRRQVQGH	-EVDPDWBAATIAOMQLKNILLMFQEW	-----ACDE	557			
Db	505							
Db	657	SLLTACKLFGAWKIKRGEHVLHEDQNFISYLEYTSIYIQTAEKVSEKSDIDS	716					
Qy	558	ELLIVAYKECHKAVMRCSTSESSSKTVQSCGSHLETKSYRVEDSLVTHPLSRTL	617					
Db	717	KLFLNA	-----IRIISFFIGNRSYKLIYDSHEVIRFVSHERVAFMNPLOTMLSF	768				
Qy	618	LHVRLSRGLGASRLH	-----EFVSFEDFQVEVLVEYPLRCLVLVAQWAEWMRRNGLSLSIQ	674				
Db	769	LIEKVSUKDAYEALDECDFLKISDFS	-----LRSVLCSDIDVGFWRNGMSVLHQ	820				
Qy	675	VFYQDVQCREM	-----YDKDITIMQIGA	-----SLMDPNKFLLLVLQRYELAEAFNKTISTKQD	730			
Db	821	ASVY	-----KNPELGSYSRDIHNLAILAWERDDIPRIYINILDRWELLDWFTGEVDYQHT	877				
Qy	731	DLIKQNTLIEBMLQVLIYIGERYVPGVGNVKE	-----EVTMRIIHLICIEPMPHSAIA	787				
Db	878	VYEDKISFIQOQIAFIYQIUTERQYFTFSSLKDRMDQIKNSIYNLYKPLUSYKLL	937					
Qy	788	KNLPEN	-ENNETGLENVINKVATKKP	-GVSGHGVIYELKDESLKDFNMFYHYHSKYTOHKS	845			
Db	938	RSVPDYLTEDTTFDEALEEVSFVEPKGLADGVFKL	-ASL	-----YAKVDPK	987			
Qy	846	AEMQK	-----KRRQENKDEALPPPPPECPAFSKVINLLNCDIM	-----MYILRT	893			
Db	988	LLLENPESSATIIKSHLAKDKOBIAKVVLIPOVS	-----IKOLBKDALNLGAFTTNT	1041				
Qy	894	VFERAI	-----DTDSNLTMEGLMAFHILALG	-LLEEKQOLQRA	-----PEEVEFTD	940		
Db	1042	VFAKVYKLLQVCLDWEUSTFNLHLLHVHIGIFRDEDELINGKDSIPAYLSKPCINULLS	1101					
Qy	941	FYHKASRLGSSAM	-NIQMLLEKLGIPQLEGQKDMITWILQMPDTYKRLREKSKSLIVAT	998				
Db	1102	IANAKSDVSESIVRKADYLLEKM	-----IMKKPNELFESLIA	1139				
Qy	999	TSSESIKNDEITHDQK	-----AEKRKAEARLHRQIKMAQMSALQKNFIETH	1048				
Db	1140	-----SFGNOYVNDYKDKKLRQGVNLQETEKERKRRLAKHQAARLAKFNNQOQTKPMKEH	1194					
Qy	1049	KLMYDNTSMPCKEDSIMEEESTPAVSDYSRIALGPKRGPVSVEKEVLTICILQEEQEVK	1108					
Db	1195	ESEFD	-EQDNDVDV	-----GEKVSESDFTALCQDSS	-----1227			
Qy	1109	IENNAWLSACVOKSTALTUHRGKPIELSGBALDP	-----LFMDPDLA	1151				
Db	1228	-STDFFVIFA	-----YHDSPIFRPGNITFNPENFMPMDWGFYNDEKQAYIDDDVL	1277				
Qy	1152	YG-TYTGSGC	-----HYMHAVCWQKYFEAVQLSSQRIHVLDLFDLESGEYLCPLCK	1201				
Db	1278	EALKENGSCGRKVPVSCNHHIHNCFKRYVOKRFS	-----SNAPICPLCQ	1324				
Qy	1202	SLCNTVIPIIQQOKINSENADALAOULLTARWITQTVLARI	SGYNIRHAKGENPIPIFF	1361				
Db	1325	TFSNCTPL	-CQTSKANTG	-----LSLDMFESELSL	-----1355			
Qy	1362	NOQMGSTLEHSLISFGVSESIKEMWILFATTIYRIGLKVPDPDERDPRVPMLT	1321					
Db	1356	-----DTLS	-RLFKPTEENYRTINSIFSLMI	-----1381				
Qy	1322	WSTCAFTIOAIENLLGDEGKPLFGALQNRQHGLKALMQFAVAORITCPQVLI	-----Q	1375				
Db	1382	-SOCQGFRAVKRKRANFSHKVSLISVHWANTISML	-----EIASRLEKPYISIFFRSREQ	1437				
Qy	1376	KHLVRLSVLPLNKSIEDTPCLLSIDLPHVLGVGLAVAPPSLYMDPDLQVDPSPSSSSSYNH	1435					
Db	1438	KY	-----NTLKNI	-----LVCIMLFTFVIG	-----PSMEF	-EPYPQOPDTVWNO	-NQ	1478
Qy	1436	LY	-----LFHILITMAHMLQILLTVDTGLPLQAOVEDSEAHASFFFAISO	-----1482				
Db	1479	LFQYIVRSALFSPVSLQTVTEALTTSRQPLRDFLOGLSDAEQVTKLYAKAKSIGDVLK	1538					

Qy	1483	-----YTSGSICDIPGWYLVSLKNGKITPYLRCAALF---FIYLL	1520
Db	1539	VSEQMLFALRTISDVMEGLDSEIYYDLAYTELLKSL-----LPTIRRCILVFIKVLHEL	1594
Qy	1521	GVTPEELHTNSAEGEYSALCSVLSLPTNLFLFLFOEYWDVTVRPLLORWCADPALLNCLKQ	1580
Db	1595	KDSENETLVINGHEVEE-----LEFDTAEFVNKALKMITEKESLVDLUTT	1641
Qy	1581	KNTVYPRKRN-----SLIEPLDDYSCLLNOASHPRCPRSAD-----DERKHPV	1625
Db	1642	QESIVSHPYLENIPIYECYGLIKLIDLSKLYNTVVTOSKEIKLBEERSQHKMNADNRIDFK	1701
Qy	1626	LCLFCGAIL-----CSQNICCOEIVNGEEVGACIFHALHCGAGVCIFLK	1669
Db	1702	ICLTGVKVLHRADRHMTKHLNKNCFKPGFAFLMPNSSE-----VCLHLT	1747
Qy	1670	IRECRVVLVEGKARGCAYPAPYLDEYGETD-PGLKRGNPULHSRERYRKLUHLVWQHCII	1728
Db	1748	QPPSNIPI-----SAPYLSNHGEVGRNAMRRGDLTTLNKKRYEHLNRLWINNEIP	1797
Qy	1729	BEIARSOE-----TNQMLEGFN	1745
Db	1798	GYISRVMGDEFVRTILNSGLFAFN	1822
RESULT	4		
UBRL_KLULA			
ID	UBRL_KLULA	STANDARD;	PRT; 1941 AA.
AC	O60014;		
DT	15-DEC-1998	(Rel. 37, Created)	
DT	15-DEC-1998	(Rel. 37, Last sequence update)	
DT	15-DEC-1998	(Rel. 37, Last annotation update)	
DE	N-END-RECOGNIZING PROTEIN (UBIQUITIN-PROTEIN LIGASE E3 COMPONENT) (N-RECOGNIN)		
DE	GN	UBRL	
OS	Eukaryomyces lactis (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.		
OX	NCBI_TaxID=28985;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Waller P.R.H., Varshavsky A.;		
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.		
CC	!- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS		
CC	TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING		
CC	ACCORDING TO THE N-END RULE, BUT DOES NOT BIND TO OTHERWISE		
CC	IDENTICAL PROTEINS BEARING STABILIZING AMINO-TERMINAL RESIDUES.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for commer-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/annou		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
EMBL;	AF061554; AAC15841.1;		
DR	InterPro; IPR003126; Znf_UBRL1.		
DR	InterPro; IPR001841; Znf_ring.		
DR	Pfam; PF02207; zf-UBRL1; 1.		
DR	SMART; SM00184; RING; 1.		
DR	SMART; SM00396; Znf_UBRL1; 1.		
DR	Ligase; Ubiquitin conjugation		
SW	SEQUENCE 1941 AA; 232682 MW; 37C2E1BCA0803268 CRC64;		

Query Match 7.0%; Score 647.5; DB 1; Length 1941;

Best Local Similarity 20.5%; Pred. No. 7.6e-32;
Matches 409; Conservative 319; Mismatches 689; Indels 581; Gaps 91

QY 35 FYTAFLLHHLAQIVPEIYYFAEMDPDLKEQESVQMISFTPLEWYLFGEPPDICLEKLKHS 94
::: :||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 51 YYYNYMSDSGLPLHMFTATNREPRPKKNVDQAMETKLSKPKWKIDENG--HSKFNIAG 108
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

Db 1397 SSLRKQYDEEKCELLQVQDLSFKVDTLSKEKISALEQVDDWS 1439

```

RESULT 6
MYSA_CHICK
ID MY5A_CHICK STANDARD; PRT; 1829 AA.
OC Q02440;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN VA (MYOSIN 5A) (DILUTE MYOSIN HEAVY CHAIN, NON-MUSCLE) (MYOSIN
GN HEAVY CHAIN P190) (MYOSIN-V).
MYOSA.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93012002; PubMed=1383040;
RA Espresafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
RA De Camilli P.V., Larson R.E., Mooseker M.S.;
RT "Primary structure and cellular localization of chicken brain
RT myosin-V (p190), an unconventional myosin with calmodulin light
RT chains.";
RL J. Cell Biol. 119:1541-1557(1992).
CC -!- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC FORMATION (BY SIMILARITY).
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -!- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES
CC (POTENTIAL).
CC -!- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X67251; CAA47673.1; -
CC EMBL; Z11718; CAA77782.1; -
CC HSSP; P08799; 1MND.
CC InterPro: IPR002710; DIL.
CC InterPro: IPR000048; IQ.
CC InterPro: IPR001609; myosin_head.
CC Pfam; PF01843; DIL; 1.
CC Pfam; PF00612; IQ; 6.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; MYOSINHEAVY.
CC ProDom; PD003376; DIL; 1.
CC SMART; SM00015; IQ; 6.
CC SMART; SM00242; MYSC; 1.

```

```

DR PROSITE; PS50096; IQ; 6.
KW Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Phosphorylation.
FT DOMAIN 1 766 MYOSIN HEAD-LIKE.
FT DOMAIN 767 789 IQ 1.
FT DOMAIN 790 814 IQ 2.
FT DOMAIN 815 837 IQ 3.
FT DOMAIN 838 862 IQ 4.
FT DOMAIN 863 887 IQ 5.
FT DOMAIN 888 915 IQ 6.
FT DOMAIN 916 1239 COILED COIL (POTENTIAL).
FT DOMAIN 1315 1419 COILED COIL (POTENTIAL).
FT DOMAIN 1661 1766 DILUTE.
FT NP_BIND 163 170 ATP (BY SIMILARITY).
FT DOMAIN 644 666 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1734 1734 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 1142 1142 E -> EQ (IN REF. 2).
SQ SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

Query Match 2.0%; Score 180.5; DB 1; Length 1829;
Best Local Similarity 17.7%; Pred. No. 0.0054;
Matches 312; Conservative 269; Mismatches 633; Indels 545; Gaps 85;

QY 94 GAFOLCGRVFKSGTETYSRCD---CAI-----DPTCVLC-----MDCFQDSVHKHRYKMH 142
DB 324 GIFRILAGILHLGNVEFASRSDSCAIPPKHDLPTIFCDLMGVD-YEEMAHWLCHRLAT 382
QY 143 STGGFCDCGDTAEAWKTPFCVNHPEGRAGTIKENSRCPLNEEVIVQARKIPSPVIK--- 199
DB 383 A-----TETY-IKPSIKLHAINARDALAKHIYANLFWIVDHVKNKALHSIVKQHS 431
QY 200 YVEMTWEEKEKLPPELQIREKNERYCVLFNDE--HHSYDHYIVYLSQALDCELAEEA 257
DB 432 FIGVLDIVGPE-----TPEINSFEQ--PCINYANEKLOQQFNHMFVKLEQE---EYMKBP 481
QY 258 LHTTAIDKEGR-----AVKAGAYAAQCEAKEDIKSHSENVSOHPHVEVLHSEIMAHQ 311
DB 482 IPWTLFDYDQPCINLIEAKMGVLDLDECKMPKSGDDTWAQKLYNTHL-----N 533
QY 312 KFAL-----RLGSMWNK--IMSYSSDFRIFQACALREPEPDSNPCLISRLMLWDKLYKG 365
DB 534 KALFEKPLRS--NKAFIIKHFAKVEYQCEGFEKKNDT-----VVEE 575
QY 366 ARKILHELIFSPFMEYKKLFAMEFYKQLOKEVISDDHDSISITALS-----VOM 421
DB 576 QIKVL-----KSSKKFKLLPE--LFQDEEKAISPTSATPSGRVPL 613
QY 422 FTVP-----TLARHL-----IEQNVISVITETLEVLPEYLD--RNKNFPGYS 465
DB 614 SRTPVKPAKARPGOTSKEHKKTGVGHOFNLSHLLMETLNATTPHYVRCIKPNDFEPFTF 673
QY 466 QDKLG---RVYAVICDLKYILISKPTIWRERLMOFLEGRFSFLK---IL-----TCMQ 513
DB 674 DEKRAVQOLRACGVLETIRISAAGFSPRWYQ---EFFSRVRLMKQKVDLSDRKQTCN 730
QY 514 GMEBI-----RRQVG-----OHIEVDPDWEAAATAIQMLKNILLMFQEWCA 554
DB 731 VLEKILDKDKYQFGKTKIFFRAGOVAYLEKIRADKLRAACIRIQTIRGWLMRKKYMRM 790
QY 555 CDELLLVAYKECHKAVMRCSTSFISSKT--VVOCSGHS-LETKSYRVSEDLVSIHLPL 611
DB 791 RRAAITIQRVYVRGHQA--RCYATEFLRRTRAATIIQKFORVMYVVRKRYQCMRDAT---IAL 845
QY 612 SRTLAGLHVRLSRLGAVSRHLEFVSFEDQVEVLVEYPLRCLVLVAOVVAEWRNGLSL 671
DB 846 QALLRGYLVLR--NKYQMLRHKSI-----IIQKHVRCW----- 877
QY 672 ISOVFYQDVKRC-----REEMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAEAFNK 723
DB 878 LARVHYHRTLKAIYVYLQCCYRRMMAKRELKCLKTEARSVERYKKLHGLE-----NK 929
QY 724 --TISTKDQDLIKOYNLTIEEM--LQVLIYIVGERYVPGV-----GNVTKEEVTM 769

```

```
Db 930 IMQORIDEQONKEYSLLEKMMNLEITYTETEKLRSDVERLRMSDEEAKNATNRVLSL 989
Qy 770 REIHLICIEPMPSALAKNLPENNETGLEINVINKVATFKPGVSGHGVYELKDESL- 828
Db 990 QEEI-----AKLRKELHOTOTEKKYIEWADYKHETELVS-----ELKEONTL 1034
Qy 829 -----KDFNMVYHYKSTQSHAKHMQK-----RRKQ-----ENKDBALPPPPPECFPAFSK 877
Db 1035 LKTEKEELNRRIHDOAK-----EITETMEKKLVEETKQLELDLNDERL-----R 1078
Qy 878 VINLLN-----CDIMYILRTVFERAIDTDSNLTWTEGMLQMAFHILA--- 919
Db 1079 YONLLNEFSRLERYDDLKDEMNLVMSIPKPGHKRTDTHSS--NESEYTFSSITEAED 1136
Qy 920 LGLEEQOQLOKAPEEVETDFVHKASRLGSSAMNIOMLLEKLGIPQLEGOKDMITWIL 979
Db 1137 LPLRMEPEKKAP-----LDMSEFLKLRQVTELEOEROSLO--- 1174
Qy 980 QMFTVRLREKSLIVATTSGSESIKNDITHDKAEAKRRKRAEAARLHRQKIMQMSA 1039
Db 1175 --DELDRKEQALRAKAKEERPPIRGALEYESLK--ROELSES--ENKLLKNELE 1225
Qy 1040 LQKNFIETHKLYDNTSEMPGKEDSINEEESTPAVSYSRIALGPKRGPSVTEKEVLTCI 1099
Db 1226 LQKALTETR-----APEVTAPGAPAY-RVLL-----DQUTSV 1256
Qy 1100 LCQEQEVKLENNAMVLSACVQKSTALTOHRGKPIELSGALDPLFMDPDLAYGTYTGSC 1159
Db 1257 --SEELEKVEELILRSQVLSQKEAI-----QPKDKNTMTSTILLEVDQVMKDKGEI 1309
Qy 1160 GHVMHVMQKYPEAVQLSQQRHVDLFDLEGEYLCPLCKSL-----CNTVPIPI 1212
Db 1310 AQAVIGLKTNRLLLESQLOKSKSHENELESRLGE-----IQSLKEENRQOQLAONLQ 1364
Qy 1213 LQPO-----KINSENADALAQL-----LTLARWITQTVLARISGYNIRHAKG 1253
Db 1365 LPPEARIEASLOHEITRLTNELDLMEQLEKQDKTVRKLLKQKLVFAKKIGLEVQGMEN 1424
Qy 1254 ENPIPFNGMGDSL-----EFHSILSFGVSESIK----- 1285
Db 1425 ISPGQII-----DEPIRPVNIPIRKEKDFQGMLEYKKEDEQKLVNILELKPGRGAVNL 1478
Qy 1286 -----YNSIKEMWILFATTVIRIGLKVPPDERPRVPMLTW--STCA 1326
Db 1479 IPGLPAVILFCVVRHADYLDNDQKVRSLTSTINGI-KKVLKRGDDFETVSWLSNCR 1537
Qy 1327 FTTOAIENLGDGKPLFGALQNRHNGKALMQFAVA---QRITCPQVLIQRIHLVRLS 1383
Db 1538 F-LHCLAKQXGEBGFMKHNTPRQNEH---CLTNFDLAERYQVLSDLAIQIYQQLVR--- 1589
Qy 1384 VLPNKSIEDTPCLLSIDLFLHVLGAVLAPSLYWDPDVQLPSSVSSSYNHLFLHLIT 1443
Db 1590 -VLENI-----LQPMIVSGMLEH----- 1606
Qy 1444 MAHMLQILLTV-DTGLPLAQVOEDSEAHSSAFBISQYTSIGSCDIPGVWLSLK 1502
Db 1607 --ETIQGVSKVFTGL-----RKRTSSIADEGYTTLDSTIIRQNSRHS-VMCQ 1651
Qy 1503 NGITPYL-RCAALFFHYLLGVTPPEELHTNSABGEYSALGSY---LSLPTNLFLLFOEYW 1558
Db 1652 HGMDPELIKQVVKQMYIIGAVTLNLLLR-----KDMCSWSKGMQIRYNV----- 1697
Qy 1559 DTVRPLLQRKADPALLNCLKQNTKQNTVVYRPRKNSLIELPDDYSCLLNQASHFCPRSAD 1618
Db 1698 -----SQLEEWLRDKNLMNS-GAKETL-----EPLTQAQLLQVKKKTD 1735
Qy 1619 DERKHPVLCFLCGAILCSQ 1637
Db 1736 EDAAE-AICSMCNALTAQ 1752
```

RESULT 7

```
REST_HUMAN
ID REST_HUMAN STANDARD; PRT: 1427 AA.
AC P30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RESTIN (CYTOPLASMIC LINKER PROTEIN-170 ALPHA-2) (CLIP-170) (REED-
DE STERNBERG INTERMEDIATE FILAMENT ASSOCIATED PROTEIN).
GN RSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood monocytes;
RA MEDLINE=92405160; PubMed=1356075;
RA Bilbe G., Delabie J., Brueggen J., Richener H., Asselbergs F.A.M.,
RA Cerletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,
RA de Wolf-peeters C., Shipman R.;
RT "Restin: a novel intermediate filament-associated protein highly
RT expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL EMBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=92405160; PubMed=1356075;
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RA "CLIP-170 links endocytic vesicles to microtubules.";
RL Cell 70:887-900(1992).
CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64838; CAA46050.1; -
DR EMBL; M97501; AAA35693.1; -
DR PIR; S22695; S22695.
DR MI; 179838; -
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01302; CAP_GLY_2.
DR SMART; SM00343; ZnF_C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
KW Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.
FT DOMAIN 60 125
FT DOMAIN 143 204
FT DOMAIN 214 279
FT DOMAIN 304 331
FT DOMAIN 350 1342
FT DOMAIN 1408 1421
FT VARSPIC 457 491
FT CONFLICT 1069 1069
FT CONFLICT D -> E (IN REF. 2).
SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;
Query Match 1.9%; Score 175.5; DB 1; Length 1427;
Best Local Similarity 19.4%; Pred. NO. 0.0077;
Matches 200; Conservative 160; Mismatches 344; Indels 329; Gaps 45;
QY 167 EPGRAGTIKENSRCPLNEEVIVQARKIPFSVIKYVVVETIWEELPPELO-----IREK 222
```

```
Db 488 EDTRVATVSEKSR-----IMLEKDLALRVOEVAELRRR 521
QY 223 NERYCVLFNDEHHYDHYIYLSQALDC-----ELAAQLHTTAIDKEGRRAVAKAGA 275
Db 522 LESNKPAGDVMSLLQEISSLQEKLEVTRTDQREITSLSKEHFGAREETHQKEKA-L 580
QY 276 YAACO---EAKEDIKS-----HSENVSOQHLVHEVHSEIMAHQFAURLG-SWNKNIMS 326
Db 581 YTATEKLSKENEKSLKLEHANKENSTALWKSRLKLETAIAHQOAMEELKYVFSKGLGT 640
QY 327 YSSDRQJIFCO-----ACLRPEPDSENCLISRLMLWDKAKLYKGARKIILHELI 374
Db 641 ETAEPFAELKQLEKRLDYOHIEINLQNOODSER----- 674
QY 375 PSSFFMEMEYKLFAMEFYKQKQKQKEYSIDDDHRSISITALSQVMTPTPLARHLIEE 434
Db 675 -AAHAKEMALRAKLMKYIK-EKENSLEAIRSKLOKAE-----OHLVEM 717
QY 435 QNVISVITET-----LLEVLPYLDNRNKF--NFQGY---SODKLGRYVAVTCDLKYILI 484
Db 718 EDTLNKLOEAEIKVKELEVLQAKCNEQTKVIDNFTSQLKATEEKL-----LDL----- 765
QY 485 SKPTIWTBRLMQLFEGPSFLKILTCMOGMEIIRVOYCOHLEVPDMEA--AIALQML 542
Db 766 -----DALRKASSEG-KSEMKKL--ROOLEAABEKQI-KHLEIKENAEKSSITREL 814
QY 543 KNILLMFQWACDBELLIVAYKECHKAVMRCSTFISSSKTVWGSCGSHLSKTSYRVSE 602
Db 815 QG-----RELKLTNLOE-----NLSEVSQVKETLEKELQILKEFAEASE 854
QY 603 DLVSTHLPRLSRTLAGHVLRLSRLGAVSRHLEFVSEDFQVEVLVYPLRCLVLVAQVVAE 662
Db 855 EAVSVQSRMQETVNLHQKEEQFNMLSS-----DLEKL----- 887
QY 663 MWRRNGLSLSIOVFYOD-----VKCREMYDKDI--IMLOIGASLMDPNKFLLLVLQR 714
Db 888 ---RENLDMEAKFREKDRBEOLIKAKEKL-ENDIAELMKNGSNGSSOLT----- 935
QY 715 YELAEAFNKTISTKQDDILKQYNTLIEEMQLVLIYVGVGVNVTKEBVTMWREIHH 774
Db 936 -----MDELRLKERDV-----BELQLKLTANEN--ASFLQKSIEDMTVK----- 974
QY 775 LCIEMPHPHSAITAKNLPENNENETGLENVINKVATFKPGVSGHGYELKDESLEKDFNNY 834
Db 975 ---ABQSOQEAAKKHEEKKELERKLSDEKKMET-----SHNQOQELKAR----- 1017
QY 835 FYHYSKTOHSAEHMOKKRRKROENKDEALPPPPPPFCFAPSKVINLNCNDIMMYILRTV 894
Db 1018 --YERATSETKTKEE-----ILQNL 1036
QY 895 FERAIDTDSNLWTECMLQMAFHILALGLLEKQKQLOKAPAEVEVTFDYHKASRLGSSAMN 954
Db 1037 OKTLTDTEK--KGAREE-----NSGLLQLEULELRKQADKA-----KAAQTAEADAMQ 1082
QY 955 I--QMLEKLGKIPLEGOKDMITWLOFDTVKRLRKESCLIVATTSGSESIKNDEITH 1012
Db 1083 IMEQWTEKETETLASIEDTKQNAKLQNLDELDTLKENNLKNVEELNKSRELLTIVENQKWE 1142
QY 1013 DKEAERKKAFAARLHRQKIMQAQNSALQKNFTIETHKLMYDNTSEMPGKEDSI-----ME 1067
Db 1143 FRKETETLKQAAQK-----SQQLSALQE-----ENVKL-----AEELGRSRDEVTSHQKLE 1189
QY 1068 EESTAVSDYSRIALGPKRGPSTVEKVLTCILCQEEQEVKTIENAMVLSACV-QKSTAL 1126
Db 1190 EER-----SVLNNQLLEMKKRKSFKIDA-----DEEKASLOKKSITSITALLTEKDAEL 1238
QY 1127 TOHRGKPIELSGE 1139
Db 1239 EKLNRNEVTVLGE 1251
```

RESULT 8

```

MYS2_SCHPO
ID MYS2_SCHPO STANDARD; PRT: 1526 AA.
AC Q9USI6; P78969;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN TYPE II HEAVY CHAIN 1.
GN MYO2 OR SPC645.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=972;
RX MEDLINE=98075862; PubMed=9415380;
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;
RT "Type II myosin involved in cytokinesis in the fission yeast,
Schizosaccharomyces pombe.";
RL Cell Motil. Cytoskeleton 38:385-396(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE
CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.
CC MAY WORK IN CONJUNCTION WITH MYO3.
CC -!- SUBUNIT: BINDS TO CDC4 AND RLC1.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75357; AAC49908.1;
CC EMBL; AL049496; CAB39901.1;
CC HSSP; P08799; LMND.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ; 1.
KW Myosin; Actin-binding; ATP-binding; Coiled coil; Calmodulin-binding;
KW Alkylation.
FT DOMAIN 1 757 MYOSIN HEAD-LIKE.
FT DOMAIN 758 787 IQ.
FT DOMAIN 875 1244 COILED COIL (POTENTIAL).
FT NP_BIND 170 177 ATP (POTENTIAL).
FT DOMAIN 634 656 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 734 748 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 674 674 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1337 1337 S -> R (IN REF. 1).
SQ SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;

Query Match 1.9%; Score 175.5; DB 1; Length 1526;
Best Local Similarity 19.9%; Pred. No. 0.0085;
Matches 230; Conservative 147; Mismatches 425; Indels 351; Gaps 54;

QY 39 FLHHLAQLVPEIYFAEMDPLEKQESVQMSITPLEWYLFGE--PDI-CLEKLKHSGA 95
Db 473 FNHHMFVLEQEEYM-----KEIV-----WDFIDFGHDLQPTIDLEKANPDI 516
```

Qy 96 FOLGRVFKSGTETYSRCD--C-----AIDPTCVLCMDFQDSVHKHNRHRYKMHSTGGF- 148
Db 517 L-----SCLDECVMPKATDAFTSKLDALR--NKSLKYKPKFKADQGF 560
Qy 149 -----CDG-DTEAWKTPFCVNEHPEGRAGTIKNSRCPLNEEVIVQARKIPPSVIKYV 202
Db 561 LTHYAADVPYSTEGW-----LEKNTDPLNENV---AKLLAQSTNKRVA 600
Qy 203 EM-----TIWEEKELPPEL--OIREKNERY--CVLENDEH-- 235
Db 601 TLFSDYQETETKTVGRTRTKGLFTVQARHKEQLNQLMNQNSTQPHTRICVINEEKKM 660
Qy 236 HSYDHVYLSQALDCELAELAAQTHTTADKEGRRAVAGAYAAQCAQEKEDIKSHSENVSO 295
Db 661 HTPNR-----PLVLGQLRCNGV-LEGIRITRAGF-----PNR 691
Qy 296 HPLHVEVLHSEIMAHQFALRLGSMNMKIM-----SYSDFRFOIFCOACLREPPD 345
Db 692 LPFNDFRVRVEIMAHLPDGTGVESRRASVMIELBKIDEASVYRGVSKIFFKAGVLAEL 751
Qy 346 SENPCLISRLMLDAKLYKGARKTLHELIFSSFFMEME-----YKLFAMEFVK 394
Db 752 ERRVATLQRLM---TMLQTRIRGELQRIKIFQKRLKDQAIKLQANLQVYNEFRTPFWAK 808
Qy 395 YKQLOKEYISDDHDSRISIT-----ALSVQMFVTPPTLARHLIEONVISVITETL 445
Db 809 LFFNLRLPLLSTQNDKQLKRDABIEILKVELKKQOONSKEVEDLVTNLSLAV-ENL 867
Qy 446 L-----EVLPEYLD--NNKFNFGYSQDKLGRVYAVICDLKYLISKPTIW 490
Db 868 LTTAIALDKKEILTRTOERLANIEDSFSETKQOENLQRESA---SLKQI----- 916
Qy 491 TERLMOFLEGRFLKILTCMQGMEIRROVGQH-----IEVDPDWEAAIAIQMLKML 546
Db 917 NNELESELLE---KTSKVETLLSQNELKELKLEEKDLDTKGELESRLNENATVLSK 973
Qy 547 LMFQEWACADELLVAYKECHKAVMRCSTSFSSSTTVQSCGHSILTKSVRYVSEDLVS 606
Db 974 AEFNEQCKSQIETVTKDAELDKL---TKYISDYKTEIO---EMRLTNOKNKSQ 1024
Qy 607 IHLPLSRTLGLHVLRLGAV-SRLHEFVSFDFQVEVLVEYPLRLCIVLVAQVVAEMR 665
Db 1025 QEGSLSESLK---RVKLERENSTLSDVSILKQOKEEL-----SVLKGQVQELTI 1071
Qy 666 RNLGLISQVFIYQ-DVVC-----REMYDKDIIMLOIGASLMDPNKFLLLVLQRYEL 717
Db 1072 NN---LEEKVNYLEADYKQLPKLKELESNDKD-----OLYOL 1107
Qy 718 AEFNKTIKTDOLIKOYNTLIBEMLOVLIYVGERVYVPGVGNVTKKEVTMRETIHLIC 777
Db 1108 QATNKELEAKVKCLNNIKSLTKELN-----KEKQCNLSASLKYIELQEL----- 1156
Qy 778 IEPHPSAIAKNLPENNETGLENVINKVATFKKPGVGHGVYELKDESLKDFNMVYFH 837
Db 1157 ---H-----ENLLKVSLENKKYKVEGL-QLDLEGLKQVDTNQE 1193
Qy 838 YSKTQ-----HSAKHEMOK-KRRQENKD-----EALPPPPPEFCAPFSKVINL 881
Db 1194 LSKKHRLDTFNHESLLRQSASVYKEKLSSENKDLNKSLSLTKQVNELSPKASKVPKL 1253
Qy 882 LNCDDIMVILRTVPERAIDTDSNLWTEGMLQMAFHILALGLLEEKQLOKAPEEVTFDF 941
Db 1254 -----ERKI-----TNLMHYSQJGKTFDEKRRAL 1279
Qy 942 YHKASRLGSSAMNTQMLEKLGIPQLEGKDMITWILQMDPTVKRLREKSKCLIVATTS 1001
Db 1280 I--ASRNEELRSLLKSELSKRKL-EVEYOK-----VLEEVKTTLSRSEVTLRNKQVAD 1331
Qy 1002 SESIKNDIEHDKKAEKRRKAEARLHROKIMAQMSALQKNFTIETHKLYMDNTSEMPGK 1061
Db 1332 HESIRS-KLSEVENKLVDRK-----ELNSALDSCKKREAEIHRLEKRRPS---GK 1378

Qy 1062 EDSIMEESTPAV 1074
Db 1379 ENNIPAVKTTTPV 1391
RESULT 9
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Sziliak L., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis".
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor".
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1".
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15005; CAA78727.1; -
DR PIR; S28261; S28261.
DR HSSP; P03069; LZII.
DR MIN; 117143; -
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335
FT COILED COIL (POTENTIAL).
FT DOMAIN 336 2471
FT GLOBULAR (POTENTIAL).
FT NP_BIND 2472 2663
FT ATP (BY SIMILARITY). 93

Db	1451	LEHHGLGRAHEEAENOPDLSREWORA.NFOOAISALQYVQPHTUTSGLLVSAVVVRGLOPA	1510
Qy	688	YDKDI-----IMLQIGASLMDPNKFTLLVLQRLEYELAEAFN	722
Db	1511	YGYGHPHAWWSLVTHSLPYFGSKSLGTVPFVVQICKNLDD-----LVKQYE-SESVK	1562
Ov	723	KTIST--KDODILIKOYN-TLIREMLQVLIIVCGERVYPGVGNVTKEFVTWMTIRIHLICIE	779

Db	1563	LSVSTRKRNISPDYPTLLGLELTIHFCLLEQ----	ANONKKTMAAGDPANL----	1613	
Qy	780	PMPSAIAKNNIPENNNETGLENINKVATFFKPGVSGHGVYELKDESKDFNMFYFHYHS	839		
Db	1614	NRARNAILEELPRTVTNTMALLNNVIRKEETQKRP-----	VOLLGATKGSSSVYF-KTT	1665	
Qy	840	KTOHSK-----	AEHMQKRRKQENKDEALPPPPPEF-----	CP	873
Db	1666	KTIROKILDFNLPLTAHLGVOLTAAVAAYWSRKKQAQRKSMKIIPTASASQLTVDLVC-	1724		
Qy	874	AFSKVINLLNCDIMMYILRTVTER-----	AIDTDSNLTGEGMLQMAPHIL-----	918	
Db	1725	-----ALSTLOTDTLLHKEVVKRPQVKGGEKSPLDVPIVGLQFYAFQLRPLVPALQE	1780		
Qy	919	-----ALGLEEKQOOLQKAPEEVTFDYHKASRLGSSAMNIQMLLEKLGIPOLEGOKDM	974		
Db	1781	NFSSLLGVLKESVOLNIAPP-----	GVFTLLSMINDFVTRTPNLENNKKDO	1825	

QY 9/5 ITWLQMFDTVKRLREKSLIVATTSGSEKNDETHDKERAERKRRAEAARLHQKIM 1030

Db 1826 -----KDLQEITQKILEAVGNIAGSS-----L 1847

Qy	1033	ADMSALQANFIDTHAKNDNISMPKEDP-----SIMEEESIPAVSDSKRLALGAFRG	106
Db	1848	EQTSWLSRN-LEVKAQOPQASLEESDAEEDLYDAAAASAMVSSAPSVYSQVALSL-----	1901
Qy	1088	PSVTEKEVLTCL-----CQEEQEVKQIEN-----NAMVLSACVQKSTA	1125
Db	1902	---LAEVLASLDMVYSDEKEKAVPLISRLLYVFPYPLRNHSAYNAPSFRAGAQLLSS	1957
Qy	1126	LTQHRGKPIELSGEALDPLFMPDPDLAYGTYTSCGHVHMHAVCQKYFEAVOLSSQQRHIV	1185
Db	1958	LSGYAYTKRAWKKEVLE-LFLDP--AFFOMDTS CVH-----WKSIIIDL-LTHEKTMFK	2007
Qy	1186	DLFDLESG-----EYLCPLCKSLCNPVIP-----IIPLOPKIN-----SENAD	1224
Db	2008	DLMNNQSSSLKLFSSFEQKAMLLKROAFVSGELDQYHLYLPLIOERLTDNLNRVGQTSI	2067
Qy	1225	ALAQLLTARWJOTVILARSIGNIRHAKENPIPI-----FFNQMGDSTLEFHSILSFG	1279
Db	2068	VAAQMFLEFR---VILLRISP---QHLSLWPIMVSELITQTFQLEEDLKDEDESURSTN	2121
Qy	1280	VESSIKYS-----NSIKB-----MVLIFATTIYRIGLKVPPDRPRVPMLTWSTCAF	1327
Db	2122	KVNRTKVSPDANGFSVGEIPOSELILYLISACKFLDTALSFPD-----KMPLFQIYRWAF	2177
Qy	1328	TQATENLLGDGKPLFGALQNRQHNGLKALMQFAVAQRITCPOVLQIKHLVRLLSVW--	1385
Db	2178	IPEV-----DTEGPAFTSDVBEENHOECK-----PHTVRIELELLK	2212
Qy	1386	-LPNIKSEDTPCLLSIDLPHVLVGAVLAFPSLYWDDPVDLPQSSVSSSYNHLVFLHLITM	1444
Db	2213	KFGEISSDSEITMKS-----EEPL-----LRQHSVSSIROLMPPFFMTLNG	2252
Qy	1445	AHMLQILTLTVDTP-GLP	1459
Db	2253	AFKTRQLPADSPGTP	2268

RESULT 11

AKA9_HUMAN

ID AKA9_HUMAN

STANDARD:

PRT: 3911 AA.

AC Q95996; Q9U004; Q9UQH3; Q9Y6Y2; O14869; O43355; O94895; Q9Y688;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

RESULT	11
AKA9	HUMAN
ID	AKA9_HUMAN
AC	Q99996; Q9U0Q4; Q9U0H3; Q9Y6Y2; O14869; O43355; O94895; Q9Y6B8; PRT; 3911 AA.
DT	20-AUG-2001 (Rel. 40, Created)
DR	20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE A KINASE ANCHOR PROTEIN 9 (PROTEIN KINASE A ANCHORING PROTEIN 9)

DE (PKA9) (A-KINASE ANCHOR PROTEIN 450 KDA) (AKAP 450) (A-KINASE ANCHOR

DE PROTEIN 350 KDA) (AKAP 350) (HGAKAP 350) (AKAP 120 LIKE PROTEIN)

DE (HYPERION PROTEIN) (YOTIAO PROTEIN) (CENTROSOME- AND GOLGI-LOCALIZED

DE PKN-ASSOCIATED PROTEIN) (CG-NAP).

GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 4).

RC TISSUE=Brain;

RX MEDLINE=98151389; PubMed=9482789;

RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;

RT "Yotiao, a novel protein of neuromuscular junction and brain that

RT interacts with specific splice variants of NMDA receptor subunit

RT NR1.1";

RL J. Neurosci. 18:2017-2027(1998).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.

RX MEDLINE=99219864; PubMed=10202149;

RA Witzak O., Skalihegg B.S., Kerker G., Bornens M., Tasken K.;

RA Jansen T., Oerstavik S.;

RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring

RT protein located in the centrosome, AKAP450.1";

RL EMBO J. 18:1858-1868(1999).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 3).

RC TISSUE=Brain;

RX MEDLINE=99287934; PubMed=10358086;

RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;

RT "Characterization of a novel giant scaffolding protein, CG-NAP, that

RT anchors multiple signaling enzymes to centrosome and the golgi

RT apparatus";

RL J. Biol. Chem. 274:17267-17274(1999).

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RA Kemmer W.A., Deiss S., Schwarz U.;

RT "Cloning of Hyperion.1";

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).

RX MEDLINE=99115654; PubMed=9915845;

RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.;

RA Trotter K.W., Milgram S.L., Goldenring J.R.;

RT "AKAP350, a multiply spliced protein kinase A-anchoring protein

RT associated with centrosomes.1";

RL J. Biol. Chem. 274:3055-3066(1999).

RN [6]

RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).

RC TISSUE=Lymphoblast;

RA Hinds K., Sutterer C., Becker M., Hawkins M.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).

RC TISSUE=Lung;

RA Milgram S.L., Goldenring J.R., Schmidt P.H.;

RT "AKAP350: A multiply spliced family of proteins with centrosomal

RT association.1";

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [8]

RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).

RC TISSUE=Brain;

RX MEDLINE=99087487; PubMed=9872452;

RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.;

RA Tanaka A., Kotani H., Nomura N., Ohata O.;

RT "Prediction of the coding sequences of unidentified human genes. XI.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro.1";

RL DNA Res. 5:277-286(1998).

RN [9]

RP SEQUENCE OF 17-1800 FROM N.A.

RA Wu X., Graves T., Bradshaw H.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE

CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND

CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL

CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN

CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-

CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR

CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS

CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.

CC -!- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N

CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)

CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.

CC -!- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND

CC CYTOPLASMIC IN PARIETAL CELLS.

CC -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3/CG-

CC NAP, 4/YOTIAO, 5 AND 6/AKAP350; ARE PRODUCED BY ALTERNATIVE

CC SPLICING.

CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY

CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.

CC -!- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,

CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A

CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.

CC -!- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO

CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.

CC -!- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR

CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AJ131693; CAB40713.1; -

CC EMBL: AB019691; BAA78718.1; -

CC EMBL: AJ010770; CAA09361.1; -

CC EMBL: AF026245; AAB86384.1; -

CC EMBL: AF083037; AAD22767.1; -

CC EMBL: AC004013; AAB96867.1; ALT_FRAME.

CC EMBL: AF091711; AAD39719.1; -

CC EMBL: AB018346; BAA34523.1; -

CC EMBL: AC000066; AAC60380.1; ALT_FRAME.

CC MIM: 604001; -

CC Coiled coil; Alternative splicing; Polymorphism.

KW DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.

FT DOMAIN 164 914 COILED COIL (POTENTIAL).

FT DOMAIN 944 1022 COILED COIL (POTENTIAL).

FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).

FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).

FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).

FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).

FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).

FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).

FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).

FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).

FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).

FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).

FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).

FT DOMAIN 3726 3730 POLY-LIU.

FT DOMAIN 203 292 GLN-RICH.

FT DOMAIN 321 1010 GLU-RICH.

FT DOMAIN 1846 2772 GLU-RICH.

FT VARSPLIC 17 28 MISSING (IN ISOFORM 2 AND ISOFORM 3).

FT VARSPLIC 1637 1642 OLOEEI -> LATRD (IN ISOFORM 4).

FT VARSPLIC 1643 3911 MISSING (IN ISOFORM 4).

FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 3).

FT VARSPLIC 2175 2183 SADTFQKVE -> Q (IN ISOFORM 6).

FT VARSPLIC 2895 2907 VGFYNNCFSTLC -> GSSIPELAHSDAYQTREICSS

```
FT VARSPLIC 2895 2948 (IN ISOFORM 2, ISOFORM 3 AND ISOFORM 6).
FT VARSPLIC 3901 3911 MISSING (IN ISOFORM 5).
FT STQFHAGMR -> ALSITTSQWHSARPATPLFFILSH
FT SLG (IN ISOFORM 6).
FT K -> KO.
FT /FTid-VAR_010926.
FT E -> Q (IN REF. 3).
FT M -> I (IN REF. 3).
FT E -> S (IN REF. 3).
FT R -> G (IN REF. 3).
FT N -> S (IN REF. 3).
FT H -> N (IN REF. 3).
FT K -> N (IN REF. 3).
FT QKH -> PKP (IN REF. 1 AND 2).
FT Q -> P (IN REF. 1 AND 2).
FT N -> P (IN REF. 1 AND 2).
FT V -> D (IN REF. 3).
FT V -> E (IN REF. 3).
FT R -> P (IN REF. 1 AND 2).
FT N -> T (IN REF. 3).
FT V -> G (IN REF. 3).
FT MISSING (IN REF. 5).
FT A -> P (IN REF. 3).
FT I -> V (IN REF. 3).
FT V -> D (IN REF. 5).
FT EI -> HE (IN REF. 7).
FT E -> V (IN REF. 3).
FT L -> R (IN REF. 3).
FT I -> N (IN REF. 8).
FT E -> D (IN REF. 3).
FT P -> S (IN REF. 3, 7 AND 8).
FT Q -> H (IN REF. 3).
FT Q -> H (IN REF. 3).
FT ESE -> OSO (IN REF. 3).
FT P -> A (IN REF. 3).
FT T -> S (IN REF. 3).
FT 3911 AA; 453664 MW; 3FB1CB1C819B47AA CRC64;
SQ SEQUENCE
```

Query Match 1.8%; Score 168.5; DB 1; Length 3911;
Best Local Similarity 18.7%; Pred. No. 0.09;
Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;

```
QY 203 EMTIWEKEKLPPELOIREKNERVYCVLFNDEHSHYDHIYVLSQALDCELAQAOLHTTA 262
DB 728 EMTL-QINLOKEIETLROEK-----EKGLEQEVQELQ-LKTELLEKOMK--- 772
QY 263 IDKGRRAVAGAYAAQCAEKEDIKSHSNVSOHPLHVEVLHSEIMAHQKFAIRLGSMWN 322
DB 773 -EKENDLOERFAQLEAENSILKDEKKTLEDMLK-----IHTPVQEEERLI-----FLD 819
QY 323 KIMSYSSDF---RQIFCOACLRPEPDSNCLISRLMLWDAKLYKGARKILHELIFSFF 379
DB 820 SIKSKSDSVWEKEI--EILIEENEDLKQOCI-----OLNDEEKEKORTFSAEKN 868
QY 380 MEMEYKKLFAFMEYKVKYKOLQKEY-----ISDDHRSISITALSVQMTPTLIARHLIEE 434
DB 869 FEVN-----YQLEQEBYACLLKVKDDLEDKSN-----KQLEY 901
QY 435 QNVISVITETLLEVLPEYLDNRN-----KFNFGYSQDKGRVYAVICDLKYLILISKPTIW 490
DB 902 KSKLKALNEEL-----HLQRINPTTVKMKSSVFDEDK--TFVAETLEMGEVVEKDTTEL 953
QY 491 TERLRMQFLEGRFSLKILTCMGMREIRQVQCG-HIEVDPDWEAATAIOMQLKNILLMF 549
DB 954 MEKLEVTREK-----LBSQLSDLSQKQKHGEISFLNEEVKSLQKEQVSLRC 1006
QY 550 QEW-----CACDPE-----LLLVAYKECHKVAMRCSTSFSSSKTVVQSCG 590
DB 1007 RELEIINHNRANVOSCDTQVSSLLDGVVMTISRGAEBSVKVKNKSGFGESEKIMVED-- 1064
QY 591 HSLETKSVRYSED-----LVSIHLPLSRTLGLH-VRLSRLGAVSRLHEFVSVFEDFOVEV 644
```

```
DB 1065 -KVSFENNTVGEESKQEQOLIIDLHLPVTKESLSRATQPSNDKLOKELNVLSEQNDLRL 1123
QY 645 LVEYPLRCLVLVAQVVAEMWRNGLSLISOVFYVDVKCREEMYDKDIIIMLOIGASIMDP 704
DB 1124 QMEAGRICLSLVS-----THVDQVREYME-----NEKDKAICLSKEELIFA 1165
QY 705 NKFLLLVLQR-YELAEAFNKTIKTDDQ-----LIKQYNTLIE-----MLQVLIYIVGR 754
DB 1166 QEEKIKELQKIHOLELQTMKTQETGDEGPLHLLIKGLQKAVSECSYFLQTLCSVLGEY 1225
QY 755 YVPGVGNVTKEEVTREIHLILCIEPMPSHATAKNLPENNETGLENVINKVATKKPG 814
DB 1226 YTPAL-----KCEV-----NAEDKEN----- 1241
QY 815 VSGHGVYELKDESLDENFYHYVSKTOHSAEHMOKKRRKOENKDEALPPPPPPFCPA 874
DB 1242 -SGDYISENDEPELODYRYEVDQFOENMHT-----LLNKVTEEYNK----- 1281
QY 875 FSKVINLLNCIDIMYILRTVFERAIDTDSNLW---TEGMLQMAFHIALGLLLEKQOOLQ 931
DB 1282 -----LLVLQTRL-----SKIWQOOTDGM-----KLEFGEE-----N 1308
QY 932 APEEEVTF-----DFYHKASRLGSSANNIOMLEKLGIPOLEGOKDMITWI 978
DB 1309 LPKEETEFLSHSQMTNLEDIDVNHK-----SKLSSLODLEKTKLEQVQELSLISS 1362
QY 979 LOMEDTVKRLREKS-CL-----IVATTSGSESINDEITHDKEKAERKKAARLHRQ 1031
DB 1363 QOOLKETEQNYEAELHCLQKRAVSESTVPPSLPVDVSVITTESDA-----O 1409
QY 1032 KIMQMSALQKNFIETHKLWYDNTSEMPGKEDSIMEEESTPAVSYSRIALGPKRGPSVT 1091
DB 1410 RTMYPGSCVKKNI-----DGTIEFSG-EFGVKEETNI-----VKLL 1444
QY 1092 KEVLTCTLCOEEOEQEVKTIENNAMVLSACVOKSTALTQHRGKPIELSGEALDPLFMDPLA 1151
DB 1445 EKQY-----QEQLBEVAKVIVSMIAFAQOQTELSR-----ISG----- 1478
QY 1152 YGTYTGSGHVMHAYCWOK--YFEAVOLSSQORIHVDLF---DLESGEYLCPLCKSL 1203
DB 1479 -GKENTASSKQAHAVCQEQEHYFNEMKL-SQDQIGFQTFETVDVVKFEBFKPLSKREL 1533
RESULT 12
BP28_DROME
ID BP28_DROME STANDARD; PRT: 2096 AA.
AC Q9VW75;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN CGI10805.
GN CGI10805.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arkil J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
```


de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasner K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaldali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*:"
RL Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO THE BAP28 FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HEAT REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF003615; AAF52447.2; -
DR FlyBase: FBgn0031864; CG10805.
DR PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.
KW Hypothetical protein.
FT REPEAT 2058 2094
FT SEQUENCE 2096 AA; 237217 MW; 3E7B3C67CE6DF62C CRC64;
Query Match 1.8%; Score 164; DB 1; Length 2096;
Best Local Similarity 18.7%; Pred. No. 0.069;
Matches 289; Conservative 221; Mismatches 611; Indels 428; Gaps 73;
QY 181 PLNEEVIVQARKIFPSVIKYVEMTWIWEKEKLPPELQIREKNERYCYVLFNDEHSYDH 240
DB 181 PLAKTAINRAASNPALFGICOST-OKAVKELGPRAHQLOAQINFYATV----- 221
QY 241 VIYSIORALDCELAELAAQLHTTAIDKEGRAV-----KAGAYAAQCEAKEDIKSHS--- 290
DB 222 VVGALQATAPLQ-----DWHITITLESLLRLGSLTSDNIDFMAAYVIVAQLVSRTKLKSVC 277
QY 291 -----ENVSOPLHVEVLHSE-----IMAHOKFAL-----RLG-SWMNKMISYS 328
DB 278 NALLERVANCPP-----ERLHSESLLLIVCIYKQQAALPHKPKETILNLVCKKWLITLSSL 335
QY 329 SDFP---QIFCQ-----ACLRPEPSDNPCLISRLMLWDARLYKGARKILHELIFSSF 378
DB 336 AKGNIAIOSICMPLMTGAVAAIRDDASSNSC-----KLFDLNL 374
QY 379 FPMEMYKLFAMEFVYKYKQLOKEYITSDHDSISITALSVMQFTVPTLARHLIEQNV 438
DB 375 LSEVPMPKPTAQOQLNCFLDITYE-----TAIDAP-----EPMETNSNEDDDTI 418
QY 439 SVITETLLEVLPELDNRKNFQOQ-YSQ--DKLGRVYAVICDL--KYILISKPTTWITER 493
DB 419 VIDSDDEIE-----TEKTTFAQWYSTYLEKLERRYPEAFDLSVKEALRSKSTSNQ 470

QY 494 LRMOFLGFRSFLKILTCMOGMBEIRVOGQHI-EVDPDWEAAIAIQMLKNILLMFQEW 552
DB 471 KALKLALGFR-----LNTDEKAKHAYEKLYHYSADWRLS-AVOKLLQNLVNTKKR- 520
QY 553 CACDEELLVAYKECHKAVMRCSTSFSSSKTVVQSGCHSLETSYRVSDELVSI--HLP 610
DB 521 -----ERSVKLLQCELPDRINDSGAVVSTLLSLPT-----EELAEMLGFLP 562
QY 611 LSRTLAGLHVRLSRLGAVSRHFEVSEDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLS 670
DB 563 LAQTLCHEL-----LYRAQSEKDEWQPW--PLAVRHULTSALVSGSYDTNLVL 608
QY 671 LISQVFFYQDVKCEEMYDKDIIMLOIGASLMDPNKEL--LLVLQRY-----ELAEAFN 722
DB 609 LALMPLLPFGEEALAEHQH--KALRILGSDFSVKVPFALAEALKVSNKTSDFNVGHRQHFL 666
QY 723 KTISTKDDQDLIKOVTNLTIEEMLOVLIYIGERYVPGVGNVT-----KEEVTREI 772
DB 667 DIASSNOELSSOERAL-----LQSVEDHGGLYIYQKASQLTHLLLLTAYAKRELOPRES 722
QY 773 IHLCTIEPMPHSAT---AKNLPENNETGLENVINKVATFKK----PGVSGHGYELKD 825
DB 723 LHMLEKIGLYSRRLOFRVNGSQNTQNCAPLQLYVDFELLTVKNTKTALASTPWNQMTD 782
QY 826 E---SLKDFNMYFYHYSKTOHSAEHMKRRKQENKDEALPPPPPEFCFAFKVINLL 882
DB 783 ELRLCURLLEIIICAQVFESEKADQPERQEWTRALQSQSLILP-----BAQDRLEVIS 834
QY 883 NCDIMMYILRTVFERAIDTDSNLTNTEGMLQMAFHILALGLEEKQOLQKAPEEVTFDFY 942
DB 835 N----FY-----VFERL-----PELWPRDSYAVFRLOQFIILE---AVLSNPKSQIDCGLV 879
QY 943 H---KASRLGS-----SAMNIQMLLEKLGIPQLEGQDKDMITWILQMFDTVKRLREKS 992
DB 880 HVLRVANACGSPLQTLRVQAINILQLISNRKLVSHVE---QLVRSLLQKRSLSMDHEQY 936
QY 993 CLIVATTSGESIKNDIEITHDKEAERKRAEAARLHROKIMAOMSALQKNFIETHKLMY 1052
DB 937 ALILYITL-----EPEKATAKERLVLSKLR-SVLALASDPKQSPICITASLL- 982
QY 1053 DNTSEMPGKEDSIWEESTPAVSDYSRIALGPKRGPVTE-----KEVLTICILQEEQEV 1107
DB 983 ---AALKHVNDENFLNELLPLGLDSLKTITAGEDNQNIKQLPWPHPSEIYKSVIERFGRV 1039
QY 1108 KIENNAWLVSACVQKSTALTQHRGKPIELSGEALDPLDMPDLAYGTY----- 1155
DB 1040 -----ALNVLRLKDLAW-----KLFEDSFAQYDIYVQLEQKLOPLPC 1076
QY 1156 -----TGSCGHVMHVCWKQYF-EAVOLSSQQRHVDLFDLESGEYLCPLCKSLCN 1205
DB 1077 VLLNSLTPETFEQMHAHKKIALIKLIVESATNSDNDISFL-----ASHRLLKKRCLDCQ 1130
QY 1206 TVIPII-----PLOPOKINSENADALQALLTLARWQ--TVLARISYNTIRHAK-- 1252
DB 1131 PLVPIILEMANTVEKKQPKRRRSVQATQD--LTSYWKQGMTLLEL-----LEHKKQL 1183
QY 1253 --GENPIPIFNQMGDSTLEFHSILSGVESSTKYSNISKEMVILFATTIYRIG---LK 1307
DB 1184 VGAELLIPPLFELLOACLATWEEHSAEYP-----KQLISLLHCCQTAQASAGVOLVK 1236
QY 1308 VPDERDPRVPMLTWSTCAFTIOAIENLLGDGPKPLGALQNRHNGLKALMQFAVAORI 1367
DB 1237 AMPES-----SPRIELVQSLNRTRNP-----QTQOH-----ALL-FLTHCAG 1273
QY 1368 TCQOVLIQKHLVRLLSVVLNPKSEDTPCLLSIDLFLVLCVAVLAFPSLWDDPDVLOPS 1427
DB 1274 MYPQQQLHK-IVEIFTVGSTVARHDD--AFSLHIHNVVESII-----PILL--- 1318
QY 1428 SVSSSVNHLVFLHILTAHMLQILLTVDTGLPLAQVQEDSEEAHSASSFFAEISQYTSGS 1487
DB 1319 -LNTIGHNELVI-----PVLKGFADICTDPV-----HRRLLPLVATLFRVLE-- 1358
QY 1488 IGCIDIPGWYLWWSLKNGITPYLRCAALFFHYLLGVTPPEELHTNSABGEYSALCSYLSP 1547

```

Db      1359  -----PKEHLWQFL-----CIIFESQVLLQVDPQKVSTDK-----SRLDFA 1394
Qy      1548  TNLFLLFQEQYDWTVPRLQRMCA DP--ALLNCLCKOKNTVVRYPRKRNSL 1594
Db      1395  RELTLMFEE-----DPTVAIQTCIRLLDYLAKLPA TKSSL 1428

RESULT 13
MY5A_MOUSE
ID AC MY5A_MOUSE STANDARD; PRT; 1853 AA.
IC Q99104.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN VA (MYOSIN 5A) (DILUTE MYOSIN HEAVY CHAIN, NON-MUSCLE).
GN MYOSA OR DILUPE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=91141583; PubMed=1996138;
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RA "Novel myosin heavy chain encoded by murine dilute coat colour
RT locus".
RL Nature 349:709-712(1991).
RP [2]
RN REVISIONS.
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RA Nature 352:547-547(1991).
CC -!- FUNCTION: MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR
CC INVOLVED IN DENDRITE FORMATION.
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC ential entities requires a license agreement.(See http://www.isb-sib.ch/announcements
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL; X57377; CAA00651.1; -.
CC PIR; A46761; A46761.
CC HSP; P08799; 1MND.
CC MGD; MGI:105976; Myo5a.
CC InterPro; IPR002710; DfL.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF01843; DfL; 1.
CC Pfam; PF00612; IQ; 6.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC ProDom; PD003376; DfL; 1.
CC SMART; SM00015; IQ; 6.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ; 6.
CC Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Phosphorylation.
FT DOMAIN 1 765 MYOSIN HEAD-LIKE.
FT DOMAIN 766 788 IQ 1.
FT DOMAIN 789 813 IQ 2.
FT FT

```

```
QY 970 GOKDMITWILQWFDVTKRLREKSKCLIVATTSSESIKNDKDEITHDKKAEKRKAARLUH 1029
Db 1167 QEKQL-----MODELDR-----KEQVFRSKAKKEERQPIRGAEL- 1201
QY 1030 RQKIMAOMASALQKNFET-HKLMYDNTSEMPCKEDSIWEEESTPAVSDYSRIALGPKRGP 1088
Db 1202 -----EYESLKQLESENKKLNEL-----RKALSEKSNAP 1235
QY 1089 SVTE-----KEVLRCILCOEBOEYKIENNAMVLSACVOKSPALTQHRGKPIELSGE 1139
Db 1236 EVTAPGAPAYRVLMEQLTSV--SEELDVKEEVLTLRSOLVSQKEAI-----QPKDDKNT 1288
QY 1140 ALDPLFMDPDLAYGTYTSCGHHVAVCWQYFEAVQLSSQORIHVDLFDESGEYLCPL 1199
Db 1289 MTDSTILLEDQVKMKDKGEIAQAYIGLKETRLLLESQLSQKRSHEAEALRGE----- 1343
QY 1200 CKSLCNTVIPIPLQPOKINSENADALALTLTLARWIOTVLARISG---YNIRHAKGENP 1256
Db 1344 IQSL-----KEENRQQLLAQNQL-----PPEARIEASLOHEITRLTEN- 1385
QY 1257 IPIFFNOGMGSTLEFHSI-----LSFGVSESIKYSINSIKEMVILFATTI--YRI 1304
Db 1386 --LYPEELYADPKKYQYRISLYKRMIDLMQLEKQDKTVRKLLKQKVPFAKKIGELEV 1443
QY 1305 G-----LKVPDPDERD----- 1314
Db 1444 GOMENISPGQIIDEPIRPVNIPIRKEKDFQGMLEYKREDEQKLVKKNLILELKPGRGVAVNLI 1503
QY 1315 PRVPMLTWSTC-----AFTIOATENLL--GDGKPLFGALQN----- 1349
Db 1504 PCLPAYILFMCVRHADYLNDDQKVRSLTSTINSIKKVLKRGDDFETVSWFLSWTCREL 1563
QY 1350 ---RQHGLKALMQPAVAQRITPCOVLIQKHLVRLLSVVLNPKSEDTPCLLSIDLF--- 1403
Db 1564 HCLKYSGEEGEM-----KH-----NTSQNEHCLTNFDLAEYR 1597
QY 1404 HVLVG-AVLAPSLVWDDPVDLPQSSVSSYNHLYFLHITMAHMLQILLTV-DTGLPLA 1461
Db 1598 QVLSDLAIQIQQLVRVLENILQPMIVSGMLEH-----ETIQGVSGVKPTGL--- 1644
QY 1462 QVQEDSEEAHSASSFAETISQVTSIGCDIPGWVLYWSLKNGITPYL-RCAALFFHYLL 1520
Db 1645 -----RRTSIADEGTYTLDSILRLQNSFHS-VMCOHGMPELIKQVVKQMFYIV 1694
QY 1521 GVTPEELHTNSAEYSA LCSY--LSIPTNLFLLFQEWYDVTVPRLQWCADPALLNC 1577
Db 1695 GAITLNNLLR-----KDMCSWSKGMQIRNV-----SOLEEWLRDKNLMNS 1736
QY 1578 LQKQNTVYRPRKNSLIELPDYSCLLNQASHFRCPRSADDERKHPVLCFLFCGAILCSQ 1637
Db 1737 -GAKETL-----EPLIQAAQLLQVKKKTDDDAE--ATCSMCNALTAAQ 1776

RESULT 14
MY5A_HUMAN
ID MY5A_HUMAN STANDARD; PRT; 1855 AA.
AC Q9V4I1; O9UE30; Q9UE31; Q07902; Q16249; O60653;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE MYOSIN VA (MYOSIN 5A) (Last annotation update)
DE MYOSIN VA (MYOSIN 5A) (DILUTE MYOSIN HEAVY CHAIN, NON-MUSCLE)
DE MYOSIN HEAVY CHAIN 12 (MYOXIN).
GN MYO5A OR MYH12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Skin;
RA Meurers B.H., Zimmermann R., Vosberg H.P.;
RT "The complete cDNA for human myosin heavy chain 12, a class V
myosin.";
```

```
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CYS-1246.
RX MEDLINE=97351514; PubMed=9207796;
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RT "Griscelli disease maps to chromosome 15q21 and is associated with
mutations in the myosin-Va gene.";
RL Nat. Genet. 16:289-292(1997).
RN [3]
RP ERRATUM.
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RL Nat. Genet. 23:373-373(1999).
RN [4]
RP SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).
RC TISSUE=Fetal brain;
RX MEDLINE=94245227; PubMed=8188282;
RA Engle L.J., Kennett R.H.;
RT "Cloning, analysis, and chromosomal localization of myoxin (MYH12),
the human homologue to the mouse dilute gene.";
RL Genomics 19:407-416(1994).
RN [5]
RP SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=95136715; PubMed=7835087;
RA Moore K.J., Testa J.R., Francke U., Milatovich A., Copeland N.G.,
RA Jenkins N.A.;
RT "Cloning and regional assignment of the human myosin heavy chain 12
(MYH12) gene to chromosome band 15q21.";
RL Cytogenet. Cell Genet. 69:53-58(1995).
RN [6]
RP SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).
RA Edgar A.J., Bennett J.P.;
RT "Inhibition of dendrite formation in melanocytes transiently
transfected with antisense DNA to myosin V.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP FUNCTION.
RX MEDLINE=99376094; PubMed=10448864;
RA Mehta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,
RA Cheney R.E.;
RT "Myosin-V is a processive actin-based motor.";
RL Nature 400:590-593(1999).
CC -!- FUNCTION: PROGRESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
FORMATION.
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: DEFECTS IN MYO5A ARE A CAUSE OF GRISCELLI SYNDROME (GS).
GS IS A RARE AUTOSOMAL RECESSIVE DISORDER THAT RESULTS IN
PIGMENTARY DILUTION OF THE SKIN AND HAIR, THE PRESENCE OF LARGE
CLUMPS OF PIGMENT IN HAIR SHAFTS, AND AN ACCUMULATION OF
MELANOSOMES IN MELANOCYTES. MOST PATIENTS ALSO DEVELOP AN
UNCONTROLLED T LYMPHOCYTE AND MACROPHAGE ACTIVATION SYNDROME,
KNOWN AS HEMOPHAGOCYTIC SYNDROME, LEADING TO DEATH IN THE ABSENCE
OF BONE MARROW TRANSPLANTATION.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```


Search completed: February 28, 2002, 10:06:39
Job time: 275 sec

```

Db 1075 VKEVVKENLEWVRAAQAALRQMEDAARRKO---AEEAVAKQARIEDLERAISVE 1131
Qy 576 TS-FISSKTVQSCGHSLETKSYR-VSEDLVSIHLPLSRTLAGLHVRLS-----RLG 626
Db 1132 PKVIVKEVKVEODPGLLQESSRLSLLEERTKNATLARELSDLHSKYSVVEKQRPVQ 1191
Qy 627 AVSRLHEFVSFEDFQVEVLVEYPLRCLVLVAQVVAEMRRNG-----LSLI 672
Db 1192 LQERVHEI-----FOVDPETEQEITRLKAKLOEMA--GKRSQVEKEVEKLLPDLEVLRAQ 1244
Qy 673 SOVEFYQDV-----KCREEMVXDIIIMQIGASLMDPNKF--LLVLQO-----RYE 716
Db 1245 KPTVEYKEVTQEVVRHERSPSEVLREIDRLKAQLNELVNSHGRSQEQQLIRLQGERDEWRRE 1304
Qy 717 LAEAFNKTIS-----TKQODLIKOYNTLJEMLQV-----LIY-----IVGERY 755
Db 1305 RAKVEYKTVSKEVVRHEKDPVLEKEAERLQEVREAAQKRAAEDAVYELOSRLLELRR 1364
Qy 756 VPGVGNVTKEEVPMREIHLICIEPM---PHSAIAKNLPENENNENETGLENVINKVATFK 812
Db 1365 KP-----EKVVVQEVV-VTKDPKPLREHSRLSGSLDEEVGRRRQLELEVQOL---R 1413
Qy 813 PGVSGHGVYELKDESKDFNMVYHYSHKTOHSAEHMOKARRKOENKDEALPPPPPEFC 872
Db 1414 AGVE-----EOEGLLSF-----QEDRSKKLAVERELQTLRIQIELEKRRP--- 1454
Qy 873 PAFSKVINLNCDIMMVIILRTVFERAI-DPDSNLWTEGMLQMAFHIALGALLEKQOLQK 931
Db 1455 -----TVOEKIIMEEVVKLEKDPDLEKSTEALRWDLDOEKTOVTE 1494
Qy 932 APEE-----EVTDFYHKASRLGSSAMNIQMLEKLGIPOLEGOKDMITWILQMFDTVKR 987
Db 1495 LNRECKNLQVQIDVLQAK---SQEKTYKEVIRVQKDRVLEDERARVWENLNPRTAQ 1551
Qy 988 LREKSC-----LIVATTSG-SESINDEITHDKKAER---RKAAEARLHROKIMQAM 1037
Db 1552 ARBEEARRLERIDRAETLGRWTSRESELOARQADQOCGRLOQELRALEROK---QQ 1608
Qy 1038 SALQKNFIETHKLMYDNTSEMPCKEDSIMEEESTPAVSDYSRIALGPKRGPSTVEKEVLT 1097
Db 1609 QTLQLO--EESKLLSOKT-----ESEROKAA-----ORGQELSRLE--A 1643
Qy 1098 CILCOERQEVKIENNAMVLSACVQKS--TALTQHRGKPIELSGEALDP---LPMDDPLAY 1152
Db 1644 AILREXDOIVEKERTLRLHAKYSRELSQETQOTRETNLSTKISILEPETCKDMSPEAY 1703
Qy 1153 -----GTY-----TGSCGHVMHVCWQKYFEAVQLSSQORIHVDLFDL 1190
Db 1704 KRGIIDRGVYLOLQLECEWEEVTTSGPGE-----ESV-----LLDR 1741
Qy 1191 ESCEYLCPLCKSLCNTVPIPILOPOKINSENADALAQLLTLARWQTVLARISGYNIRH 1250
Db 1742 KSGKQYSIEALRC-----RRISKEE-----YHL-Y 1766
Qy 1251 AKGENPIPIFNQMGD-----STLEFHSILSGFVSESSIKYSNSIKEMVILFATTIYRIGL 1306
Db 1767 KDGHLPISEFALLVAGETKFPSSLSIGSIIS-----1797
Qy 1307 KVPDDRPRVPMLTWSTCAFTIQAENLIGDECKPLFGALQNRHNGKALMQFAVAQR 1366
Db 1798 -----KSPLASAPQSTFFS-PSFSLGLGDDSPFIAGIYDTTDN--KCSIKTAVAKN 1848
Qy 1367 ITCP---QVLIQKH-----LVRLLSVVLNKSIEDTPCLLSIDLPHVLGVAVLAFPSLYW 1418
Db 1849 MLDPIITGOKLLEAQAATGGIVDILLSRERYSVHKAMERGLIENTSTQORLLNAQKFTGI-- 1906
Qy 1419 DDPVDLQPSVSSVSSYN-----HLYLPHL 1441
Db 1907 EDPVTKRLSVGEAVOKGWMWPRESVLPHLQVQHL 1940

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 09:58:03 ; Search time 28.88 Seconds
(without alignments)
4613.203 Million cell updates/sec

Title: US-09-724-126A-2
Perfect score: 9224
Sequence: 1 MADEAGGTERMEISAEPLQ.....ETARSQETNQLFGFNWQLL 1749
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	8617	93.4	1757	T14318	ubiquitin-protein
2	1945.5	21.1	1827	T25604	hypothetical prote
3	1038	11.3	1958	T39808	hypothetical prote
4	914	9.9	2052	T37711	probable n-end-rec
5	689.5	7.5	1950	S12332	ubiquitin--protein
6	647.5	7.0	1941	T30554	ubiquitin-protein
7	606.5	6.6	1225	T48251	ubiquitin-protein
8	549.5	6.0	1872	S64851	probable membrane
9	343	3.7	795	T48252	ecceiferum3 (CER3)
10	218	2.4	271	T40238	364K Golgi complex
11	205.5	2.2	3187	JC56337	giantin - human
12	190.5	2.1	3259	A56539	early endosome ant
13	186	2.0	1410	A57013	centrosome associa
14	182	2.0	2442	T08621	hypothetical prote
15	180.5	2.0	1413	T26467	myosin-v - chicken
16	180	2.0	1830	S19188	Tpr homolog - frui
17	179.5	1.9	2346	T13829	giantin - human
18	176	1.9	3225	I52300	restin - human
19	175.5	1.9	1427	S22695	myosin ii - flssio
20	175.5	1.9	1526	T41522	centromere protein
21	175.5	1.9	2663	S28261	microtubule-vesicl
22	174.5	1.9	1392	A43336	hypothetical prote
23	172.5	1.9	1780	T17272	probable nuclear p
24	168	1.8	1837	T41023	probable centromer
25	166	1.8	1676	E71410	hypothetical prote
26	164.5	1.8	181	T01799	kinesin-related pr
27	164	1.8	2954	T14156	restriction modifi
28	163.5	1.8	1339	H81307	myosin heavy chain
29	163	1.8	1853	A46761	

30 163 1.8 2649 2 A40937
31 162 1.8 1269 2 F84730
32 157.5 1.7 1642 2 T08880
33 156.5 1.7 1496 2 T05634
34 156.5 1.7 1790 2 S67593
35 156.5 1.7 2710 2 A37052
36 156.5 1.7 2829 2 A42771
37 155.5 1.7 1216 2 A28821
38 155 1.7 1025 2 E86355
39 155 1.7 1173 1 A53430
40 155 1.7 1828 2 B59254
41 154.5 1.7 2269 2 T28677
42 154 1.7 3685 1 A27605
43 153.5 1.7 1178 2 S30431
44 153.5 1.7 1313 2 F96673
45 153.5 1.7 1375 2 T37672

ALIGNMENTS

RESULT 1

T14318
ubiquitin-protein ligase E3-alpha - mouse
N:Alternate names: N-recognin E3-alpha
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14318
R:Kwon, Y.T.; Reiss, Y.; Fried, V.A.; Hershko, A.; Yoon, J.K.; Gonda, D.K.; Sangan, P.
Proc. Natl. Acad. Sci. U.S.A. 95, 7898-7903, 1998
A:Title: The mouse and human genes encoding the recognition component of the N-end ru
A:Reference number: Z17977; MUID:98318583
A:Accession: T14318
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-1757 <KWO>
A:Cross-references: EMBL:AF061555; NID:g3170886; PID:g3170887; PIDN:AAC40165.1
C:Genetics:
A:Gene: Ubr1
A:Map position: 2

Query Match 93.4%; Score 8617; DB 2; Length 1757;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1619; Conservative 64; Mismatches 66; Indels 8; Gaps 3;

Qy 1 MADEAGGTERMEISAEPLQTPORLASWDDQVDFYTAFLHLHAQLVPEIYFAEMDPDL 60
Db 1 MADEMDGAERMDVSPPEPLAPQRPASWDDQVDFYTAFLHLHAQLVPEIYFAEMDPDL 60
Qy 61 KOESVQMSIFTPLEWYLFGEPPDICLEKLHSGAFQLCGRVFKSGETTYSCRDCAIDPT 120
Db 61 KOESVQMSIFTPLEWYLFGEPPDICLEKLHSGAFQLCGRVFKSGETTYSCRDCAIDPT 120
Qy 121 CVLCHDCFDQSDVHKHRYKMHHTSTGGGFCDCGDTGTEAWKTGPFVCHNHEPGRAGTIKENSRC 180
Db 121 CVLCHDCFDQSSVHKHRYKMHHTSTGGGFCDCGDTGTEAWKTGPFVCHNHEPGRAGTIKESLHC 180
Qy 181 PLNEEVIVQARKIFPSVIKYVEMTITWEEKELPPELQIREKNERYCYVLFNDEHHSYDH 240
Db 181 PLNEEVIAQARRIFPSVIKYIVEMTITWEEKELPPELQIREKNERYCYVLFNDEHHSYDH 240
Qy 241 VIYSLQALDCLAEALHTTAIDREGRVAKAGATAACQAEKEDIKSHSENYSQHPLHV 300
Db 241 VIYSLQALDCLAEALHTTAIDREGRVAKAGATAACQAEKEDIKSHSENYSQHPLHV 300
Qy 301 EVLHSEINAHQKFAIRLGSGWMNKIMSYSSDFRQIFCOACLRERPDSENPCILSRMLMWA 360
Db 301 EVLHSEVMAHQKFAIRLGSGWMNKIMSYSSDFRQIFCOACLRERPDSENPCILSRMLMWA 360
Qy 361 KLYGKARKILHELIFSSFFMEMEYKKLFAFMEFYKQLOKEYISDDHRSISITALSQ 420
Db 361 KLYGKARKILHELIFSSFFMEMEYKKLFAFMEFYKQLOKEYISDDHRSISITALSQ 420

367	Db	ENILLQDTQWKRAGRSILHQMILMRVFMVYIYDQKVRFKAFKAPMLHMYNEIYEDPIDKHENDV	426
413	Qy	SITALSVQMFVPTPTLARHLIEBQNVISVITETLLEVLPEYLDNRN-----KFNFGQYS-Q	466
427	Db	SVVGLSVQMFVPSLARKLVAEDQAFSVISKRAIRDOTDKFYKYNDGKIARFDFTSRSP	486
467	Qy	DKGRVAVICDLKYILISKP--TIWTRERLQMFLGPRSEFLKILTQCMQWEERIRVOVG	524
487	Db	PELRSLHTIRDMAYTILNAPVSESDNRELIDGFGVQGFADFLLELQHLQGMDEVKQRAVE	546
525	Qy	HIEVDPDWEAATAIQMOLKNILLMPOEACACDEEL---LLVAYKECHKAVMRCSTSFSS	581
547	Db	HQWSEWETAENILRLKDAISMIIGNAETNEVHNKLMIM---CUELMNRMPVYTKS	603
582	Qy	S-----KTVQVQSGHSLKTSYRVSEDLVSIHLPSLRTLAGL-----HV	620
604	Db	EEDTYELTVTINGESCRISHFDVLKSSTSVHQPVVRIIAGLFSASNTYGFLLNRSNNSHT	663
621	Qy	RLSRGASVRLHEFVSFEDQVEVLVEYPLRCLRLVAQVVAEMRRNGLSLISQVFIYQD	680
664	Db	SLNQ---ERIKELINCKD--ETNLYELSLRVLVLCQASNATLWRNGFSLQINHVES	717
681	Qy	VKCREEMDKDIIMIQIGASLMDPNKFLLLVLQRYEL-----AEAFNKTIISTKQ	730
718	Db	PLCRNEMFDRDVLMOVGAAATPSTFKTHLLHRFLFKWATSEFEODKANEKPAKPESE	777
731	Qy	DLIKOYNLTIEEMLOVLIYIGERVPGVGNVTEEVMTREIIHLCLIEPMPISAIKANL	790
778	Db	DLSKTLMVTAEBFIOCLLILICERTYGVGKTRPMDQMKREVILHILCTGSHTSHTOOKM	837
791	Qy	PENENNE--TGLENVINKVATFKKPGVSGHGYELKDESLDFNMFYHYKSHQTSKAEHM	849
838	Db	SHDINSKRSLSEAVNLVADFCKPLATTAGOFHCKESSLPTYSPFFMHYSKDSQSAEQS	897
850	Qy	QKK--RKQENKDEALPPPPPECFAPSKVINLLNCDIMYILTVTPERAIDTDSNLWTE	908
898	Db	QARVAKMEKSVRACAPPILOFOTFERIPELITNVLIIHVRLIIDRTA--RSRFSDD	956
909	Qy	GMLOMAFHIALGLLEEKQOLQKAPBEEVTFDFYHKASRLGSSAMTOMLEKLGKIPQL	968
957	Db	RLPHKTYLIGIALNEE-----EKNPSGFTQRA-----EESIGLLSLLEGLVKP--	1002
969	Qy	EGOKDMITWLOMFDTVTKRLRKSLCIIVATTSGSES--IKNDEITHDKEAKERRKAEAA	1026
1003	Db	--ESSICPILLEV--TVKEYRK--LIKARAGVPEAPAPENKPAQSEEBIKAKARAAA	1055
1027	Qy	RLHRQKITAQMSALQKNFIETHKLWYD---NTSEMPG-----KED---SIMEES	1070
1056	Db	EM--ROKAMKMSNMOSKFM---KKIIDEKKEDESQTPSEKSETVVKXDDYDNKHFFDEDV	1111
1071	Qy	TPAVSDYSRIALGP--KRGPSVTEKEVLFTCILCOEQEV--KIENNAMVLSACVOKSTALQ	1128
1112	Db	VKQVGHDFPCIGANKWHAELVKPTLTCILCQEDDEIIAPOQGRPMVCAAFIQOSQLFT--	1170
1129	Qy	HRKGPTEL---SG--EALDPLFMDPDLAYGTYTSGSGHVMHVAWCWKYFEAVQLSGQOR	1182
1171	Db	HKNKGELMTASSGISTRDLLTAPATLQYGVDSVSTCSHSMHYECYKSLAEANRSRESLR	1230
1183	Qy	IHV-----DLFDLESVEVLCPLKSLCNTVPIIP---LOPOK-----INSENADAL	1226
1231	Db	ARQVGHSHKMYDTENGETYQCLCKRUSNAIPIVLPAYQTLNQNGFSTVSGACKENFDT--	1289
1227	Qy	AQLLTLARMIQIVLARI-----SGYNIRHAKGENPI-----PIFFNOGMG	1266
1290	Db	-----NVARVKRNLEMPLESSVSKGHSRHSRHSERSLLDLEKLDKXDPDTANTASG	1341
1267	Qy	DSTLEPHSILSGV-----ESSIKYNSIKEMVI-----	1295
1342	Db	-----VLOFGAMEMSSATHMPASBSQMLMTTSPSQDDVEFYNELAAMFVDQVNN	1393
1296	Qy	---LFATTIYRIG--LKVPPDERDPVRPM-----	1319
1394	Db	TSPAAPTETIPAIGSSRIPESQSGKKPLSQIOHVLYSLIRFPFALINSRICSSSFE	1453

RESULT 3

T39808

hypothetical protein SPBC19C7.02 - fission yeast (*Schizosaccharomyces pombe*)

C/Species: Schizosaccharomyces pombe

C/species: schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 03-Dec-1999

```
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39808
```

C;Accession: T39808

R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.

submitted to the EMBL Data Library, June 1998

A;Reference number: 221881

A;Accession: T39808
A;Accession number: 221001

A;Accession: 139808
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

A;Molecule type: DNA
 A;Residues: 1-1059 / 1000

A;Residues: 1-1958 <LYN>

A;Cross-references: EMBL:AL023859; PIDN:CAA19570.1; GSPDB:GN00067; SPDB:SPBCL

A; Experimental source: strain 972h-; cosmid c19C7

C;Genetics:

Query Match	11.38:	Score 1038:	DB 2:	Length 1958:
-------------	--------	-------------	-------	--------------

Query Match	11.3%	Score I038; DB 2; Length 1958;
Best Local Similarity	23.3%	pred No 4e-56.

Best Local Similarity 23.38; Pred. No. 4e-56;
Matches 480; Conservative 312; Mismatches 762; Tdols 502

0v 60 EKOEF5VOMSTFTPLEWY-----LFGEDPDICI,-----EKLKHSGAFOIC 99

QY 60 EKQEE SVQMSIF TPLEWY-----LFGEDPDICL-----EKLKHS GAFQLC 99

QY 100 GRVFKSGETTYSCRDCAIDPTCVLCMDCFQDSVHKNHRYKMHTSTG-GGFCDCGDEAWK 158

DB 97 GHI FRR GEFY RCKTCSVDSDNSALCVRCKFRA TSHKDHETSF T V SAGSGGCCDCGNAAWI 150

07 150 TQDECVNHEDCBACTVENCBCDI NEETVVOADKTEOCVTXVVWEMTWEE - KET DDE 216

QY 217 LQI--REKNER-----YYCVLFNDEHSHSVHYIYSIQRALDCELAEE--A 256
Db 217 ESILQDEKTSRLSENKYGIDIDSCNNYSVLWMDERKSHKQFYEQITATL--ELPNNVFG 274
QY 257 OLHTTAIDREGRRAVAGAAOAEAKEDIK--SHSENVSOHPLHVEVLHSEIMAHOKF 313
Db 275 KMANIINDIGR-----ACIVTETNLKELLKIGOKLAQINLAVSIRSMRDIPEES 325
QY 314 ALRGSMMNKIMYS-----SDRFQIFCOACLR----- 341
Db 326 CAVLLEWLADIAGSGCKGRNNYFSSVICKELVPRWNCGLHNSDLTFLRLSLRSLALPEIVA 385
QY 342 -REPD-----SENPLCLSLM----- 356
Db 386 IDSPDIFLNEDHINSSGPDSTSHMLETDESSHSRHWYPSNSLPDVLVSASVREDYFF 445
QY 357 LMDAKLYKARKILHELIFESSFEMEMEXKLFAMEFVKYKOLQKEYISDDHDSISITA 416
Db 446 LYDLKLWKSRLKQLQELYIGYFITQPGFKEIMGARITASYRRUAELUFLLLDREPEHSVIF 505
QY 417 LSVQMETVPTLARHLEEQNVISVITETLEVL-----PEYLD-----RNNKFNFOGY 464
Db 506 FSNQIFTVADVAKLLVTEYDELTTINATLYTFTYKKNLTPNYVDQHAMIRTSAAFH-- 563
QY 465 SQDKGRVAVTCDLKYILISPTIWTIERLR--MQLEGRFSFLKILTCMOGHEETIRQV 522
Db 564 -----SRRIYHIFHFIHQFMSIPCV-AEIVREDLKLQYADPFNL--PQGMCPYTRAV 614
QY 523 GQHIEVDPD-WEAAIAIQOL---KNILLMFOEW-----CACDELLLVAYPECHKAV 571
Db 615 SQHVENDSMWYLVNVSLOVAKLKHVGVMFELNTNKLANINLISILYLPKARNES 674
QY 572 MRCSTSFISSTKVQSCGHSLETSKYRVSDELVSITHLPSRLTAGLHVLRSLRGAVSRL 631
Db 675 W-TNTESLTGTIVDER-GNS-KLIEYDALQVPSEHPLHLLVYL-----L 719
QY 632 HEFVSPEDE-----QVEVL--VEYPLRCLVLVAQVVAEMRRNGLSLISOVFYQDVKCR 684
Db 720 SFYVERDNTKLLMTQDLAVTDHPLRVCAWLSQMRAKLWIRNGTTLRQDAHRYNLSFH 779
QY 685 EEMYKDIIMQIGASLMPNKLFLVLQRYELAEAF--NKTISTKDDOLIKOYNTLIEE 742
Db 780 EYTFDLDVLLQLTLTYGPDAILPSIFSRFOLEDQMYGFFVPHRHYD-VSQVTIMME 838
QY 743 MLOQLIYIVGERYVPGVGNVTKKEVTMRIIHLCTCEPMHSAIAKNLPENNETGLEN 802
Db 839 FLLLLISVCNVAVDLHDWIDTR--IEYGHAILCFRPLPYSEITKRTCEHLLHKKOFES 896
QY 803 VINKVATFKK-PGVSGHYVELKDESLKDFNMVYFYHYSKTQHSKAEHMOKKR-RKOENKD 860
Db 897 TLKKVATFRNAEGINDSGSFTLKDEYFDVYDPFNIHYSRNQREAEENILRRYSKQHSKH 956
QY 861 EALPPPPPEFCPAPSK--VINLLNCDIMMYIL-RIVFERAIDTDSNLWTEGMLQNAFHI 917
Db 957 --LESVYVEYHPILHSNITIPILQSDSFVGLWHTIVAYIYIPYDQGLGLVNTALHA 1014
QY 918 LALGLEEKQOLQKAPEEVTFDFYHKASRIGSSAMNIMLLEKLGIPOLEGOKD--- 973
Db 1015 CLLVLMSK-----GSEPIFSKKICE--NRFPPVEGLQEVCSN 1050
QY 974 ----MITWILQ-----FDTVKRLREKSLIV-----ATTSGSESINKDE 1009
Db 1051 PDVTLFSLVLCQMKNHRNFVYV--EKISLIMKLKASEVPLLYEPVYAEPLTSSSKIVQ 1107
QY 1010 ITHDKEKAERK-RKAAEARLHROKIMQAOSALQKNFIETHKLMYDNTSEMPGKEDSIMEE 1068
Db 1108 SLSDABQEQOHLAKVMAKERMARQARIMEQFRMQONKPELENHALFEASDCEN----DEADEF 1163
QY 1069 ESTPAVDSYRSLAGKPGSPSTEKVELVTCILCOBEQEVKIENNAMVLSACVOKSTALT- 1127
Db 1164 SVTSSVS--TKFLDLP-----PIDTCLLCQOEBLKKRPGYTLV---VLRSSVRLR 1209

QY 1128 -----QHRCKPIELSG---EALDPL----- 1144
Db 1210 FPADANVYSEVLDIPDSDLHDEIQERPFGLAGKRKKVLDDSTAYDYDNYEYKKGKELHQ 1269
QY 1145 -----FMDPDLAYTGTGCGHVHMAVCWOKYFEAVOLSSO---ORTHVDLFDLESGE 1194
Db 1270 LKDSFENGPPQDRLGLHATCGGFHMHDCFKNHATATVTLATRANPYRNHP--HNLMSKE 1327
QY 1195 YLCLPLCKSLCNTVPIIPILOPKINSENADALAQLTLTLARWI-----QTVL-- 1240
Db 1328 FLCPLCKALCNTIPILWRPKKEINFQEGAVLT--APLKNLWVSKTFSFNKDLNQQLDDI 1385
QY 1241 -----ARISGYNIRHAKGENPIPIFFNOGMGDSITLFEHSILSFGVSESSKYSNSIKEMV- 1294
Db 1386 ETPSEHTQSYNL-----NLLDV-LQHTLRDSLKDIYTLNTGADNS---SDNVEENAD 1434
QY 1295 ILFATTIYR-----IGLKVPDPDERPRVPMCLWTSTCAF-----TIOALENLGLGE 1339
Db 1435 NLFQSSVLDDHVFHKSVMNEVPADER-----LAISDDIFELYRRLDDVIDLNSLSYSD 1488
QY 1340 GKPLFGLQNRQHNGLKALMQFAVAQ-----RITCPQVLI-----QKHLVRL 1382
Db 1489 FIPVNGKL---HNWVK-LFSYSLCQVEASTRGHIKCSIPADIMVHNLGKNOQVFLRL 1543
QY 1383 SVVLPNKSDETPCLLSIDLFHVLVGAVALP-----PSLYWDDPV--- 1422
Db 1544 S---ESIKTYTLLC--AHDQSKRIGSGIQEIEFISFCQOKRIFGRLLPSL--DSPVTKST 1596
QY 1423 -----DLQPSVVS-----SSYNHL-YLFHLITMAHMLQIILLTVDGLPL 1460
Db 1597 TDDRVEPLLKDTREFAEASVSGLLSCDESFHYLTQLYITADIYRNLWILLSQRNSL-L 1655
QY 1461 AQOEDSEEA-----HSASSFFAEISQYTSIGSGICDI-----PG--WYLMWS 1500
Db 1656 KCMESVEFEADYEQKGFELHVIQIKWSLRVDGAGLNFDCCTEDDLNPNHLLTLYKL 1715
QY 1501 LKNIGITPYLCAALFFHYLLGVTPPEELHTNSABEYSA LCSYLSLPTNLF---LFQY 1557
Db 1716 LERFSLIFLRCALLMYCRYGVSEFTQPNLNFQNSLSRLOTQMHPGVIELSNHLCLPA 1775
QY 1558 WDTVRPLLQRWCAADPALLNCLKOKNTVVRYPKR---NSLLELPDDYSCLLNQASHFR 1613
Db 1776 SSTWLSLKHMC-----NFTETGCLCDFPRAYYPGYVELVSLPYELDKVFELLARRC 1829
QY 1614 PRSADDERKHVPVLCFGAILCSQNICQEIYNGEVGACIFHALHCGAGVCFIKIRRC 1673
Db 1830 SKCL-TEPMEPAICLFCGKLLCFQSHCCS--FNG--IGECNLHMOCASDIDGFLIVK 1884
QY 1674 RVVLVEGKARCAYPAPYLDYGETDGLKRGNPJHLRERY-RKLHLVMOQHCHIEETA 1732
Db 1885 AILYLNPPV-GSFVAPFLDAYGETDLGLRRGRSOYLSQKRYDETVRTMNLNGSIPSYIA 1943
QY 1733 RSQETNOMLFGFNQOLL 1749
Db 1944 RQLDANPDGTG--WETL 1958

RESULT 4

T37711

probable n-end-recognizing protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T37711

R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1997

A:Reference number: 221738

A:Accession: T37711

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2052 <NR>

A:Cross-references: EMBL:297208; PIDN:CA010108.1; GSPDB:GN000666; SPDB:SPAC15A10.11

A:Experimental source: strain 972h-; cosmid c15A10

C:Genetics:

A:Gene: SPDB:SPAC15A10.11
A:Map position: 1

Query Match 9.9% Score 914: DB 2: Length 2052;
Best Local Similarity 20.4%; Pred. No. 2.3e-48;
Matches 447; Conservative 326; Mismatches 734; Indels 682; Gaps 89;

QY 62 QESVQMSIETPLEW-YLFGEDP-----DICLEKLKHS-----GAFOLCGRVF 103
DB 43 QEVFLSLLNEDNWKYFLKEKPGAITSDPRLSLRLOHSEPECAQELQDKSGSKVCGHVF 102
QY 104 KSGETYSRCDCAIDPTCVLCMDCFQDSVHKHNRKYMHTSTG-GGFCDCGDEAWKTCPF 162
DB 103 RAGEVIVRCKNGCLDNTCVLCAPCFATNHEGHETHSVISYSTSGICDCGDPPEAWNVDLN 162
QY 163 CVNHEPGRAGTIKENSRCPLNEEVIIVQARKIFPSVIKVVEMTIV-----207
DB 163 CKIH-----NVPDEQKXPEVI-----PLELOHSITRTTHILLDFILDVFCSPV 209
QY 208 -----EEKELPELQIREKNERY-----YCVLFENDEHSHSDHYIS 244
DB 210 NLKASTVGSILADEAS-----RLSSAKYGVADRPCNVFRVLMWENDEVHTEDAVGS 262
QY 245 LORALDCELAQAQHTT-AIDKEGRRVAKAGA-----275
DB 263 VLEALDSNTAFGLEVAQRVDSIGRFPAVATSASVHEAIRIANAIKENLAVNVRTARDF 322
QY 276 -----YAA-----COE-----281
DB 323 REDICGILLEMFDLLESHVCYFADYLOIIVCDEILKNWSPGLEKPAKPEYFNFLPLEI 382
QY 282 -----AKEDIKSHSENV-----OHPL-----298
DB 383 VNDSDSEDDIYAAELLDVIANLODETGVTRIANGDGEDEFADMTDPTTAGFDHPLDD 442
QY 299 -----HVEVLHSEIMAHKF-----ALRLGSMNMKIMS-----328
DB 443 NDVNDLLDFETERIDIDLTDVEVMEETENEAAEADYGVNRNTRQDDVDQISMETESONE 502
QY 329 -----SDFRQIF-----335
DB 503 TDESONTENVDPQHTPVPITPTATQDVVTVIRPENSOQLNLRQIINARRPRPAAV 562
QY 336 COACLRRE-----PDS-----ENCLJLSRL-----MLWDAKLYKGARKILHELIFSSFF 379
DB 563 CQVSLSDYKWSHPPIPPSSYSFVSESSILRDYFLFDLKWKRLRGLLSKLYVVPFN 622
QY 380 MEMEYKILFAMEFYKYYKQLOKEVISDHDHRSISITALSVMQFTVPTLARHLIEEQNVIS 439
DB 623 RNLFLKRLMGIIRFVIHYRSLATAFLADRPDHSVMFLSVQFTTSPSLAEAVVKDYDFLT 682
QY 440 VITETLLEVL-----PEYLORNKNFNGYSQD-----KLRGVYAVICDLK 480
DB 683 NLNATLTLSSLTQSNRPSTL-----FSSDTEYPTIQLNRQVLKTRTRYNLFSDLG 732
QY 481 YL-----LSKPTIWTERRLMQFLGFPFSLKILTCMOGMBEIRQVGHIEVDPDWEAAI 536
DB 733 YLHQHPQVKLLVDVDRYVHQYIDLLRV-----QGVIPQQRALISHVO-----WDFPH 781
QY 537 ATOMQLKNILLMFQEWACDEEL-----LLVAYKECHKAVMRCSTSFSSSKTVVQSCGHS 592
DB 782 G-----KNILFVQWVAMLSNTVSSCFQAPYERLFAI-KCIITSITHPKLDI-----AES 832
QY 593 LETKS-----YRVEDLVSIHLPLSTLAGLHVRLSLRGAVSRLEHFEVS 636
DB 833 LEPLSCIPSSSLNTFTQPLVPFVSRODPISFYHP-----LHMMLSNLFSYCRVDASSH 885
QY 637 FEDQFQVEVLVEYPLRCLVLVAQVVAEMWRNGLSLISQVFFYQDVCKREEMWYDKDITMLQ 696
DB 886 WDKDTLLALLDHPPLRVCLVLAQIDCNLWIRNGRSILLTDFAYRLNNIEVSYDKDILAIQ 945
QY 697 IGASLMDPNKFLLLVLQRYELAE-AFNKTTISTKDQDLIKQVNTLIEEMQLVLIIVGERY 755

DB 946 TILMEVDPNVLNAVYVQRFETDWLYNLTYNEHPNYDTERIPAMLCMKLELLIALITER- 1004
QY 756 VPGVGNVTKEEVTMREIHLICIEPMHPASHIAKAKNLNPNENNETGLENVINKVATFKKP-G 814
DB 1005 -EQIUHVDIQDIIRTLAQOLCGPLAYSALLSTISSNLVESLSPKIREEVTSYKAPDG 1063
QY 815 VSGHGVYELKDESLKDFNMFYHYKSTQHSKAFHMKRRKQEN--KDEALPPPPPEFC 872
DB 1064 LHDFGVYSLKDEYDLVDOPYFYHYNKNERESDTILKRLAKKNVSAESIIEPKIRFL 1123
QY 873 -----PATSKVINLNCIDIMMILRTVFRADITDNLMTQGLM-AFHILALG--LLE 924
DB 1124 EKGDHIDFIAAVNASTFSLIIF-----RATE-----YALVQAESFGSDIGNITLG 1169
QY 925 EKQOL-----OKAPEBEVTFDYHKA-----SRLGSSAMNIQMLELKGIPQL 968
DB 1170 DALQLCLISMKIHEFSKSNDFCSRCAERYPTDSSIMREFGSGAYCLAEJCFAILKSPKY 1229
QY 969 EGOKDMITWI---LQWFD--TVKRLREKSLIVATTSGSEIKNDEITHDKAEAKRKA 1023
DB 1230 KDVHVKNVAVLAGLOKNDPDSAYSNMLEATHFELSTTSS-----TSDSNELEKTQEK 1280
QY 1024 EAARLHRO-KIMAQNSALQKNFIETHKLMYDNTSEMPGK---EDSIMEEESTPAVSYSR 1079
DB 1281 KRLALEKOKKIMOQFRDQOASFIA-----QNTDFDIGEDQTEDEVTTTEPEEVKYHEH 1334
QY 1080 TALGPKRGPSTVEKVLTCILCOEQEVEK-----IENNAMVLSACVOKSTALTQHRG 1131
DB 1335 I-----RG-----NCLLCOEENQDQAPYGVIGIIQSSLLRKTVDHSEILDEIYS 1380
QY 1132 KPIELS-----GEALDPLFMDPD-----LAY-----GTYTSCGHVMHVAWCOKY 1171
DB 1381 VPPNLDRESHSPFGKKYDTVVFNRSKORLLSAYPCGNIRGVFVSGCHLMHLGCFKNY 1440
QY 1172 FEAVOLSSQQRHIVDLFDLESG--EY-----LCPLCKSLCNTVPIIPILOPKIN 1219
DB 1441 YVARSWYRN-----DVTAGLSEYKYKSTAKFEMCPLCRSLSNVLLP-MFOIPKM-- 1489
QY 1220 SENADALAOQLTLARWIOQTVLARISYNIIRHAKGNPIPIFFNOGMDSTLEFHSI---- 1275
DB 1490 CLNIDTLNFPFRSMNGMLEEI-GTMSSSSFEYQ-----LVRSSLSDKDTFRSCFLRP 1540
QY 1276 -----LSFGVSESTIKYSN--IKEMVILFATTIYRIGLKVPPDRDPRV-PM 1319
DB 1541 WINSKIIISAMLARLIADGALLDQSNRVDSDYDRYCETT-KLAMKLVKGSFTTNNVSPH 1599
QY 1320 LTWSTCAFTTQAIENLLDGEKP-LFGALQNRHNGLKAL-MOFAVAQRITCPOVLIOKH 1377
DB 1600 DLLNSLAYTVSSLE--VSQRCSPKQSGATRSVWFNGLGPLTSLFLPLTSDT----- 1648
QY 1378 LVRLLSVLPNLIKESDTPCLLSIDLPHVLGAVLAPPSLY-----WDDPVDLQPSVV 1429
DB 1649 ---VLKVCVDQIISKDQOALLMESOKLLVCKIFYRHSOLKSLMRNGRMSDHDQIOFPFL 1705
QY 1430 SSSY-----NHLXFLHITMAHMLQILLTVDTLGLPLAQOVED---SEE 1469
DB 1706 STNFDQFVKISSMLIFGKQDNILYVVKLFYSEIKTIIISMIVKVVADSVVPLDTINYS 1765
QY 1470 AHSASSFPAELISQYTSGSIGCDIPGWYLVWSLKNGI-----TPYL 1509
DB 1766 QQSKSQFY-----ILC--KNVLLCGSSNNIEILDDESLLRLRLMSLVEKYSPLFL 1813
QY 1510 RCAAALFFHYLLGVTPP-BELHTNSAGEYSALCSYLSLPTNLLFLLQEYV-----DT 1560
DB 1814 RRVALVLYCMFDISLEFNEFSNNEDESELERLSKLIKVP-----LQELYSQMSDEENQ 1868
QY 1561 VRPLQRACADPALLNCLKQKNT-----VVYRPRKRNSLIE-LPDDYSCL 1604
DB 1869 ILELAGWCEHLA-----QNTWGDSTISLEYPGIYELVKPLHRLNLLDSMOESVCCM 1921
QY 1605 LNAQSHFRCPRSADDERKHPVLCFLCGAILC-----SONICCOEIVNGVEVGACIFHALHCG 1661

Qy	1152	YG--TYTSGCG-----HVMHVCWKQYFEAVOLSSQRIHQHVDLFDLESGEYLCPLCK	1201
Db	1278	EALKENGSGSRKVFVSCNHHIHHNCFRKYVQKKRFS-----SNAFICPLCQ	1324
Qy	1202	SLCNTVPIPIPOKQKINSENADALAQLLTLARWLTQTVLARISGYNIRHAKGENPIPIFF	1361
Db	1325	TFNSCTPLP--COTSKANTG-----LSLDMFELESL--	1355
Qy	1262	NOGMGDSLTFFHSILSFGVSESSIKYSNIKEMVILFATTIVIRIGLKVPPDERPRVPMLT	1321
Db	1356	-----DTLS--RLPKPFTTEENRTINISFLMI-----	1381
Qy	1322	WSTCAFTIAIENLLGDECKPIFGALQNRHQHGLKALMQFAQRITCPOVLI-----Q	1375
Db	1382	-SOCQGFDAVRKRAFRANFSHKDVSLLSVHWANTISML--EIASLEKPYISFFRSREQ	1437
Qy	1376	KHLVRLLSVVVLNIIKSEDTPCLLSITDLPHVLVGAFLVPPSLYWDDBDVLDLPSSVSSYNH	1435
Db	1438	KY-----KTLKNI-----LVCIMLFTFVIGK-----PSMEF-EPYQPQPDVTWNG-NQ	1478
Qy	1436	LY-----LFHLITMAHMIQILLTVDVDTGLPLAQVOEDSEEAHSASSFFAEISO-----	1482
Db	1479	LFOYIVRSALFSPVSLRQTVTEALFTFSRQFLRDFLOGLSDAEQVTKLYAKASKIGDVLK	1538
Qy	1483	-----YTSGSGICDIPGWILVWLSKNGITPYLRCAALF---PHYLL	1520
Db	1539	VSEQMLFALRTISDVMEGLDSESIYDLAYTFLLKSL---LPTIRCLVFYKVLHELIV	1594
Qy	1521	GVTPPEELHTNSAEGEYSALCSYLSLPTNLFLLFOEYWDTVRPLLQRCWADPALLNCLKQ	1580
Db	1595	KDSENETLVINGHEVEE-----LEFEDTAEFVNKALKMITESLESVLDLTT	1641
Qy	1561	KNIWVRYPKRN-----SLIELPDDYSCLLNQASHFRCPSAD-----DERKHPV	1625
Db	1642	QESIVSHPYLENIPIEYCGIITKLIDLSKYLTNTYQTSKEIKLEERSQHKMKNDRLDFK	1701
Qy	1626	LCILFCGAIL-----CSONICQCEVTNGEEVGACIFHALHCGAGVCIFLK	1669
Db	1702	ICLTGCVKVKHLRADRHMTKHLNKCFFPGAFLMPNSE-----VCLUHLT	1747
Qy	1670	IRECRVVLVEGKARGCAYPAPYLDVEYGETD--PGLKRGNPLHLSRERYKHLVWQHCII	1728
Db	1748	QPPSNIFI-----SAPYLNSHGEVGNMRRGDLTTTLNKLRYEHLNRLWINNEIP	1797
Qy	1729	BEIARSQE-----TNQMLFGFN	1745
Db	1798	GYTSRYMGDEFRVTILSNGFLFAFN	1822

RESULT 6

T30554

ubiquitin-protein ligase E3 - yeast (Kluyveromyces marxianus var. lactis)
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
C:Accession: T30554
R:Waller, P.R.H.; Varshavsky, A.
submitted to the EMBL Data Library, April 1998
A:Description: Kluyveromyces lactis UBRI1, the recognition component of the N-end rule pathway
A:Reference number: Z20856
A:Accession: T30554
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1941 <NALL>
A:Cross-references: EMBL:AF061545; NID:g3114670; PID:g3114671; PIDN:AAC15841.1
C:Genetics:
A:Gene: UBRI1
C:Superfamily: ubiquitin--protein ligase

Query Match 7.0%; Score 647.5; DB 2; Length 1941;
Best Local Similarity 20.5%; Pred. No. 8.9e-32;
Matches 409; Conservative 319; Mismatches 689; Indels 581; Gaps 91;

Qy	35	FYTAFLHHLAQLVPEIYFAEMDPDLKQESVQMSFTPLLEWYLFGEDEPDICLEKLUHSG	1
Db	51	YYYWISDSGRLLPHMFTATNEREFKNVDQAMEIKLSKKPWYKIDENG--HSKENHAG	108
Qy	95	AFQLCGRVFSGETTYSRCDAIDPTCVLCMDCFQDSVHKHRY--KMITSTGGGFCDCG	152
Db	109	--RICGAKFRGPEIYRKCKSDDTCVLCVNCNFPKDHVGHVYTSICTEFNNGICDCG	166
Qy	153	DTEAWKTPGCVNHPGCRAGCTIKENS-----CPLNEEVIQARKTFPSVIKY	200
Db	167	DKEAW-----NHELNCKGA-EDNGLEDPDDHDKGKSKMLESYLIELHDFIDFVNO	218
Qy	201	VVE--MTIWE-----EEKELPELQIREK-----NER-----	225
Db	219	NIEPLTTIQRLPLAKLRYFNPEREIYEQADMLRRLAYRNOYMEBESSNKRHLTSLDPLS	278
Qy	226	----YYCVLFNDEHHSYDVHVIYLSQRLADCELAEOHLHTTAIDKEGRRAVKAGA-YAQC	280
Db	279	TLKDYAILVYVDFHNYSQASAAIROG-GPDNKHIDLLTAKIDSEGRSLLRCSADIASLM	337
Qy	281	EAKEDIKSH--SENVSOHPLHVEVLHSEIMAHOKFALRLGWSNNKIMSY--SSDFROIFCQ	337
Db	338	GRIFSVOGNSLCTITO--WYELHQEAC--KYST--MWINDCLNPNSTFQSLFRN	388
Qy	338	A-----CLREPD-----SENPLCI-----	352
Db	389	AIGKVLCSKYEPPYQSIDMTSVVDFYSOSYLSDDPYLADHSLVGEVYKPIGLRHKSLD	448
Qy	353	-----SRU-----MLMDAKLYKGARKILHELIFSSPFMBMEYK	385
Db	449	PGDISAISPILNKVIAEDHHEYTSRLQYVFLFNRYWKKLRKIVQDLIIPTLASSAVOK	508
Qy	386	KLFAMEVKKYKOLQEYISDDHDSRISITALS---VQMTVPTLARHLIE-----Q	435
Db	509	PMFQDLVEIFPHMTRS--GTFMDREPQLTSLRESVVLQFTCTTAYSFHSGHFWYLIW	566
Qy	436	NVISVITE--TLLE--VLPEYLDORNKNFQGYSDQKLGRAVYAVICDLKYI---LISKP	487
Db	567	SVIDVDFDSTMDEGLTVWQVORSNP--SKYSISFKQGLYAVETLLSKITDPNLLKP	624
Qy	488	TIWTERLMOFLEGFRSLKILT--CMOGMEET-RRQVQOHI-EVDPDWEAAIATOMQL	542
Db	625	-----GEFIMIVTLCFLFNAGAWIKKREGEHVLREDQHFIPVLEYTTSV	669
Qy	543	KNILLMF-----QEWACADELLLVAYKECHKAVMRCSTSFISSSKTVVQSGH-----	591
Db	670	YSIIQTEDKVLQOSKDHIDQRLIGA-----INLLDSFL-----GHRNLSYK	711
Qy	592	---SLETKSVRSVEDLSYIHLPSRTLGL--HVRLS-RLGAVSRLHEVFSFEDQFQEV	645
Db	712	LYKDFEILKQISKEQVSFNNPVHTLFCFLVQHPLOVSIOVLSQSDYLVISDF-----	766
Qy	646	VEYPLRCLVLAQVVAEMRRNGLSLISQVYFYODVKCREM--YKDIIMLQIGASLMD	703
Db	767	---ALRSWVLCSDIDGFWVRNGSVLHQSAYY---KNPNEMSSYSRDIQNLQ-AFLIE	819
Qy	704	PNKFLLV--LQRYELAAEFNKITSKDODLLKQYNWLEEMQVLYIYIGERYYPGVG	760
Db	820	KNDQRYVINMLDREWLWDFDGSVPSTFTVYQDKTSIIIIQGFVAFYQLVLR-----	873
Qy	761	NVTKEEVTRE-----IHLLCTIEPMPSAIAKNLPEN-ENNETGLENIYKINVATF	810
Db	874	DFYKKFTDLEETQYLNKNAIYKLYAEPISYTDLLNDIDYLTESVSQDFTVLEEVSTY	933
Qy	811	KKPGVSGHGYELKDESLKD-----FNM--YFYHYSKTOHSAEHMQKRRKQENKDEA	862
Db	934	LEPGSKIWCFTQKEGTQLENRELRLLNMGDNDEHSATIVKSHLADSKERAKIIVKPQ-	992
Qy	863	LPPPPPPFCFAPSKVINLNCNDIMMYILTVTERAIDTDSNLWTGMLQMAHIALGL	922
Db	993	-----LLEDELDPGAR--ELGSFTRTNIFA---KLIFKLKLAV	102


```
QY 1114 -----MVLSACVOKSTALTOHGRKPIELSGEALDPLFMDPLDAYGTGSGC 1160
Db 1270 IDFTNPTDVNRINSLSFGKQKSAIQEN---PQDDGTRLFTKTCPEVLR-----ACG 1320
QY 1161 HVMHVCWQKFEAVQ-LSSQQRHVDLFDESGEYLCPCLCKSLCNTVIP-----1209
Db 1321 HGSHTKCLSGHMKSSIRGIONOTTKNIPL-SYSGGLIYCPVCNSLSNPLPKTNDIDKRTS 1379
QY 1210 ---IIPLOQKINSADALAQLLTLARWIOVLARISGYNIRHAKGNPIPIFNQOGM 1266
Db 1380 SQFFMCIEKRSEAEENLDPMSICICAAMI-----LG 1411
QY 1267 DSTLEFHSILSGVESSIKYSNIKEMVILLATTIYRIGLKVPPDERDPRVPMV--TWST 1324
Db 1412 D--LOGKKVTT--IEDAYKVNSV-----FINTISNLRSHKKEGKIVNMERISSQ 1461
QY 1325 CAFTTQATLENLGDGKPLFGALQNRHNGLKALMQFAVAQRI--TCPOVLIOKHLVRLLS 1383
Db 1462 CILTLHLVCELKSFYKKFVNS-----KTFESSELSRKINWNEFLIKGNVNL 1511
QY 1384 VVLPNI-----KSDTTPCLLSIDLH-----VLGAVLAFPSLYWDDPDVLPSPSVS 1430
Db 1512 YMSQNFNDIGGKTPPPNLCIYEMFKRRFHOLLLLARDMMRVNFYKDCRNKIKISSNG 1571
QY 1431 S---SYNHLIYFLHITMAHMLQILLTVDTGLPLAQVQEDSEAHSAFFAISIQTSGS 1487
Db 1572 SEEPSTSFYLFN--TFKKYVDLFKPPD-----VRFDTSLERIKDKICSL-----1615
QY 1488 IGDIPGWYLVSLKNGTTPYLRCAALFFHYLLGVTPPEELHTNSA-----EGEYSAL 1540
Db 1616 -----LLESISFCRFTFLFNQYDDGDDGNNRNNRNMVQKREIELI 1662
QY 1541 CSYLSLPTNLFLEQYWDTVRPLLRW-----CADPALLNCLCKQKNTVVYRPRKR- 1592
Db 1663 FRYFKLPNLTHFLKDFYFNELTQNIERYNDGNDNLRIQQVIYDMVQINTRA-YPSPEHI 1721
QY 1593 SLIELPDDYS--CLLNQASHRCPSRADDKHPVLVLCFCAILC--SONICCOEIVNGE 1648
Db 1722 QLIELPLNLSFSLDNDEISNKC-----DKYEIAVCLCGQ-KCHIQKSIALQGYLOGE 1774
QY 1649 EVGACIFHALHCGAGVC---IFLKIRECVVVLVEGKARGCAYPAPYIDYGETDPLGLK 1704
Db 1775 ---CTDH-MRNGCEITSAYGVFLMTGTGNAIYLSYK-RTGYFAAPYLSKYGETNEDYKF 1828
QY 1705 GNPLHLSRERYKHLVWQOHCIEBIARSQTNOMLEF 1743
Db 1829 GTPVYLNARYANL-----ANEIVEG 1849

RESULT 9
T48252
eCERiferum3 (CER3) - Arabidopsis thaliana
N:Alternate names: Protein Tle22.70
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48252
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224489
A:Accession: T48252
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-795 <BEV>
A:Cross-references: EMBL:AL162874
A:Experimental source: cultivar Columbia; BAC clone Tle22
C:Genetics:
A:Map position: 5
A:Introns: 203/2; 220/3; 428/1; 488/3; 666/2; 701/2; 723/3; 749/3; 772/3
A:Note: Tle22.70

Query Match 3.7%; Score 343; DB 2; Length 795;
Best Local Similarity 23.0%; Pred. No. 2.2e-13;
A:Cross-references: EMBL:AL023796; PIDN:CAA19375.1; GSPDB:GN00067; SPDB:SPBC32F12.14;
```

```
Matches 190; Conservative 123; Mismatches 274; Indels 240; Gaps 43;
QY 1033 IMQMS-----ALQKNFIETHKLMY-----DNTSEMPGKEDSIMEESTPAVSYSRTA 1081
Db 56 LMLQLSDTTISANMIESIKARLIGNQTEKRSDDGRGKDESNNM---SLEIAMYQTVR 112
QY 1082 LGPKRGSVTEKEVLTCLCOEEQEVKTENNAMVLSACVOKSTALTOHGRKPIELSG--- 1138
Db 113 -----NKIENMINQSLTRVDHQPHEAEN-----CSEKNSV-----GGSTLOGRPP 153
QY 1139 -----EALDPLFMDPLDAYGTGSGCHVHMAVCWQKFEAVQLSQQR 1183
Db 154 DIRSQRTRRPDAGSDGFHPIDCD-----GYVLSGCHAVHQSCLERLYLSKLSKERSGRT 208
QY 1184 HVD---LFDLSEGYLCPCLCKSLCNTVIPITP-----LQO-KINSENADALAQLLT 1231
Db 209 VFEGAHIVDLKKBEFLCPVCRRLANSVPECPGLCSVSKLQDSPRTKLRRKDALQPSL- 267
QY 1232 LARWIOVLARISGYNIRHA-----KGNPIPIFNQOGMDSTLEFHSILSGVESS 1283
Db 268 ---WLSALCL-----LRSAAEVIEDGDRGKTVP-----QGDGPRR-----301
QY 1284 IKYSNLSIKEMVILF-----ATTIYRIGLKVPPDERDPRVPMVLTWSTCAFTTQATLENL 1337
Db 302 -KDLKSVSKMLWDEYFPKPEDKTLKRLWL--PPQS-----IYVWDTLTKSLISME--IG 350
QY 1338 DEG-----KPLFGALONRQHNGLKALMOFAVAQR-----ITCPQVLI-QRHLVRL 1381
Db 351 TRFAKNSMLPYCIDSLYEELKTSKGTLSVLLRVQSSRTKNTIHRQRFVGMKHLAES 410
QY 1382 LSVVLPNIKSED-----TPCLLSIDL-----FHVLGAV--LAPPSLYW 1418
Db 411 ICYGVSSSSSSSIFSGEGTTGSLKNIDLLMNRASDPVLAHDPPFSLMVALFCLPPFLTC 470
QY 1419 DDVDQLQSPSSVSSYNHLYLFLHITMAHMLQILLTVDTGLPLAQVQEDSEAHSAFFA 1478
Db 471 EESL-----LSLVHIFHSVSVQTVIYACACRP-----SELSEUNFGENLLN 512
QY 1479 EISQYTSGS-----IGCDIPGWYLVSLKNGITPYLRCAALFFHYLLGVTPPE 1526
Db 513 DISNALRESGWEYFRSNMDSLCDIKD-----TIRKYSLPFLRRCALLMWLL--KSTPR 565
QY 1527 ELHTNSAEGEYSA LCSYLSLPT--NLFLFP--QYWDTVRPLLRQWCA DP--ALLNC-LK 1579
Db 566 KLH-----EESDMFDLPSPDTTNDMDFIXSPQSELNHHVQLEKMFNTPPIDIINDELL 619
QY 1580 QKNTVW-----RYPRKNS-----LIELPDDYSCLLNQASHRCRPSADDE 1620
Db 620 RSSTQIWLQHFQREYRVNRVRSKLCITPVVPFQLKLPNLQDILLQRCIKKRCV-NCTKV 678
QY 1621 RKHPVLCLFCGAILCSQ--NITCCOEIVNGEBEVGACIFHALHCGAGVCIFLKIRECRVVLV 1678
Db 679 IEEPVLCCLCGS-LCSPIWSPCCRE-----SCPNHAIITCGAGTGVFLLIRR-TTILL 729
QY 1679 EGKARGCAYPAPYIDYGETDPLGRKNPULHLSRERYKHLVWQO 1725
Db 730 QREARQSPWSPVLDTEGEEDIMIRGRKLYLNEERYAALTYLVGSH 776

RESULT 10
T40238
hypothetical protein SPBC32F12.14 - fission yeast (Schizosaccharomyces pombe) (fragme
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40238
R:Moreno, S.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, June 1998
A:Reference number: 221915
A:Accession: T40238
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-271 <MOR>
A:Cross-references: EMBL:AL023796; PIDN:CAA19375.1; GSPDB:GN00067; SPDB:SPBC32F12.14;
```


A: Experimental source: strain 972h-; cosmid c32f12

C: Genetics:

A: Gene: SPDB:SPBC32F12.14

A: Map position: 2

Query Match 2.4%; Score 218; DB 2; Length 271;

Best Local Similarity 26.0%; Pred. No. 2.8e-06;

Matches 60; Conservative 34; Mismatches 97; Indels 40; Gaps 6;

QY 60 EKQESVQMSFTTPLEWY-----LFGEDPDICL-----EKLKHSQAFQIC 99

DB 37 ESAKSLNVEFALLGVDHFLWNTLLPERPTIDASFLLRRAQGHSEGEYRHGTCESKC 96

QY 100 GRVFKSGETVYSCRDCAIDPTCVLCMDFQSDVHKNHRYKMHSTGTG-GGFCDCDGTAAWK 158

DB 97 GHIFRKEGVYRCVCSVDNSALCVKFRATSHKHDTSTFTVAGSGGCCDCDGNAAAWI 156

QY 159 TGPFVNHPEGRAGTIKENSRCPLNEEVIVQARKIFPSPVIKYVVMETIWEBE--KELPPE 216

DB 157 GDVSKIHSHEDATISNDMIDEIPEKLENSIQTTIDCVLDFVDFVSCSPENLKKMPTL 216

QY 217 LQI--REKNER-----YYCVLFNDEHSHSDHVIYSLQALD 250

DB 217 ESILQDEKTSRLSENKYGDIDDSNMYSLVWNDEKHSFKOFYQITTALE 267

RESULT 11

JC5837

364K Golgi complex-associated protein - rat

C: Species: Rattus norvegicus (Norway rat)

C: Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000

C: Accession: JC5837

R: Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.

A: Title: Identification and characterization of rat 364-kDa Golgi-associated protein re

A: Reference number: JC5837; MUID: 98093490

A: Status: nucleic acid sequence not shown

A: Molecule type: mRNA

A: Residues: 1-3187 <OK>

A: Cross-references: DDB: D25543; NID: g516825; PIDN: BAA05026.1; PID: g516826

C: Comment: This protein plays a role in the formation and maintenance of the characteris

C: Superfamily: giantin

F: 49-549, 624-1176, 1238-1707, 1763-3114/Domain: coiled-coil leucine zipper #status predict

F: 3165-3187/Domain: membrane anchor #status predicted.<MAD>

Query Match 2.2%; Score 205.5; DB 2; Length 3187;

Best Local Similarity 20.2%; Pred. No. 0.00068;

Matches 243; Conservative 182; Mismatches 472; Indels 305; Gaps 49;

QY 125 MDCFQSDVHNHRYKMTSTGGGFCDCDGTAAWKTPGFCVNHPEGRAGTIK-----ENS 178

DB 1057 IDLLQEEITENQATIQRFITMTDAGDGSAAVKETSV-----SSPPRAGGGEHWKPELEGK 1112

QY 179 RCPLEEVIVQARKIFPSVI-KYVVMETIWEBEELPELQIREKNERYCYVLFNDEHSHS 237

DB 1113 IVDLEKEKTLQKLQKALISRKAILKAAQKEKHLKXELKEQDAYRHLEQDFQDSKE 1172

QY 238 YDHVIYSLQALDCELAEAQIHTTAIDKEGRRVAVKAGAYACQAKEDIKSHSENVSQ-- 295

DB 1173 NENIRAPLRLOAKESIDQOLPGT-----GQOEPTHSEGLSLEGTPEASESDLHAAQPS 1227

QY 296 HPLHVEVLHSEIMAHQ-----KFAIRLGLSMWNKIMSYSDFRQIFCQACLREEP 344

DB 1228 HPGETATLQATVSAQIQDQLKETEVEKELEL-----KISSTTSELTKKSEEVLLLEQ 1282

QY 345 DSENPCLISRLMLWDALYKARKILHELIFSPF-FNMEYKKLFPAMEFVYIKOLQKEY 403

DB 1283 INEOGLEIQLNKAASHEAKAHTQEQLESQKADLEHLKTLQPEL-----ETLQKHV 1338

QY 404 ISDDHRSISITALSVQMTVPTLARHLIEBQNVISVITETLLEVLPEYLDNRNKNFQG 463

DB 1339 GQKEEVSYLVGQLGKEQTLTTVQTEM-BEQ-----ERLIKALHTQLEMQAKEH--- 1387

QY 464 YSQDKLGRVYAVICDLK-----YILISKPTIWTBRLRMQFLGFRSFLKILUTCQGMEEI 518

DB 1388 --EERLQVOVEICEULKQPKKELEESKAKQOLQKLOALISKEALKENKSQEQULSS 1445

QY 519 RRQVGQHIYD-PDWEAAIAIQMLKNILL-----MFOEWACADDELLIVAYKECHKAVMR 573

DB 1446 ARDAVEHLTKSLADVESQVQOEKDALGLKALLQE---ERDKLIV---ENDKSILLE 1498

QY 574 CSTSFTSSKTVVQSCGHSLETSYRVSDELVSTH---LPLSRTLAGLHVRLSR-LGAVS 629

DB 1499 -NOSLGGSCESLALGLGLTEDKE-KLMKELESVRCSKIAESTEWQKHKELOEYEVLL 1556

QY 630 RLHEFVSFDFQVEVLVEYPLRLCLIVLVAOVVAENW-----RRNGLSLISLVQFY 677

DB 1557 QSYENVSNAEIRIQHVVE-----SVRQEQEYVAKLSAESDKRERKQLODAEORME 1609

QY 678 YQDVKCRREMYDKDIIMLQI-----GA-----SLMDPNKFLLLVLQRYEL 717

DB 1610 ENKEKMRKFAKSKQKILEEBENDRLRAEAQVPVGGANESMEALLSNASLKEELERITL 1669

QY 718 AFAFNKTIISTKDQDLIKQYNTLIEMLQVLIYIVGERYVPGVGNVTKKEEVTMRRIIHLCL 777

DB 1670 E---YKTLSEPEALMAEKNTLSEETRNKLQVEAQELKQASLETTEKSDPEKDV----- 1721

QY 778 ITPMPSHATAKNLPENENNETG-LENV-INKVATFKPGVSGHGVYELKDSLDKDFNMYF 835

DB 1722 IEEVTEAVVGKSOEQDSLSNAKLEDAEATILANSARPGVSE-----TFSSHDDINNYL 1775

QY 836 YHYSKTOHSAE-HMOKRRKQ-----ENKDEALPPPPPEFCFPAFSKV-----IN 880

DB 1776 QLQDLQKGRIAELEMEKQKDRLESQTLNEKNAL----LTQISAKDSELKLEEEVAKIN 1831

QY 881 LLNCIDIMMYILRT--VFERAIDTDSNLWTEGMLQMAFHILALG----- 921

DB 1832 MLNQIQEELSRYTKLKETAEEKDDLEERLMNQLAELNGSIGNVYQDVTDAQIKNEQLE 1891

QY 922 -----LLEKQLOQKAP---EEVTFDFYHK-----ASRLGSSAMNIQML 958

DB 1892 SEMQNLKRCVSELEEEKQQLVKETKYEIRKEYMEKIQGAQKPGSKIHAKEIQ-ELL 1950

QY 959 LEKLKGIPOLEGQKDMITW---ILQMFDTVKRL----- 988

DB 1951 KEKQEVKQL--OKDCIRYLGRISALEKTVALEFVITESOKDLDAKGNLAQAVEHHK 2008

QY 989 -----REKSLIVATT-----SGSESIKND-----EI 1010

DB 2009 QAQELSSPKILLDDTQSEAAARVLADNLKLEKQSNKESIKSQIKQKDEDLRLRLEQAE 2068

QY 1011 THDKAEARKRAEABARLHRQKIMAQ-----MSALQKNFTETHKLMYDN 1054

DB 2069 KHRKEKNMQEKLDA--LHREKAHVEDTLABIQVSLTRKDKMKLEQOQSLDSTLAQAA 2126

QY 1055 TSPMPCKEDSIMEEESTPAVSDYSRIALGPKR-----GPSVTEKEVLTCLCQEBQEKI 1109

DB 2127 TKMSLSLD-----DSDRVIDEAKKWBQRFDAIQTRE-----BEVRUKE 2166

QY 1110 EN 1111

DB 2167 EN 2168

RESULT 12

A56539

giantin - human

N: Alternate names: macrogolin

C: Species: Homo sapiens (man)

C: Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999

C: Accession: A56539; S37536

R: Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.

Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (

A:Reference number: A56539; MUID:94187728
A:Accession: A56539
A:Molecule type: mRNA
A:Residues: 1-3259 <SEE>
A:Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
C:Genetics:
A:Gene: GDB:GOLGB1; GCP: GCP371
A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31
C:Superfamily: giantin
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein
F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 2.1%; Score 190.5; DB 1; Length 3259;
Best Local Similarity 18.0%; Pred. No. 0.0061;
Matches 221; Conservative 207; Mismatches 424; Indels 377; Gaps 49;

QY 164 VNHEPRAGTIKNSRCPNEEVIVQARKIPPSVIKYVVMWTWEEKELPPELQ--IRE 221
DB 12 VLHEGDDDDTDQNMRAPLDPELHQSDEMFNNTTQEDVQERLAYAE-QLVVELKDIIRQ 70
QY 222 KNERYYCVLFNDEHSHYDVHVIYSQALDCELAFAQLHT---TAIDK--EGRRAVKA 275
DB 71 KDVQL-----QKDEALQERKAADNKIKKLHAKAKLUTSLUNKYTEEMKAQGGTV 121
QY 276 YAAQCAKEDIKSHSNVSOHPLVHVEHSEIMAHOKFALRLGSMWNKIMSYSDPRQIF 335
DB 122 LPTPEQSEQLSKHDKSSFEEMEIEKIKHKLQEKELISTL----- 163
QY 336 QOACLRPEEPDENPCLISRLMLWDKLYKGARKILHELIFSSFFMEMEYKFLFAMEPVKY 395
DB 164 -QAQL-TQAQAEQAQSS-----TEME-----EFVMM 188
QY 396 YKQLOKEYISDDHRSISITALSVMQFTPTLARHLIEQN-----VISVIT 442
DB 189 KOOLQEK-----EETISTLQALSTQAEAAQVVRKDAFETQVRLHEDELQGLVT 242
QY 443 ETLEVLPEYLDNRNKNFNQYSDQKLGIVYAVICDLKYLIS---KPTIWTFLRM-- 496
DB 243 QADVETEMQOKLRVLQKLEEHESLVGR--AQVVDLQOELTAEBORNILSQOQLOME 300
QY 497 -----QFLEGRSFLKILTCMQGMEIRROVQGHIEVDPDWEAIAIQMLKNILLMF 549
DB 301 AEHTLRNATVETERESKILLKMELEVAERKLSF-----NLQEEHMHLLQEF 349
QY 550 QEWACDEEL--LLVAYKECHKAVMRCSTSF1--SSSKT--VVQSCGHSLETKSYRVSDEL 604
DB 350 EQAGQAQAELESRYSALEQKHKAEMEKEKTSHLSLQKTOGELQACDALKDQNSKLLQDK 409
QY 605 VSHILPLSLRSLAGLHVRLSLGA-----VSRL---HEF---VSFEEDQVEVLVEYPLRC 652
DB 410 NEQAVOSATQIOLEDQLOQKKEISQFLNRLPQOQHETASQTSFPDYNVNEGTAQVTEEN 469
QY 653 LVLVAQVVAEMRRNGLSLISQVYQDYQKCREMYDKDILMLQI-----GAS 700
DB 470 IASLRVVELENEKGAALLSSI-ELEELKAENEKLSQITLLAEQNRTGEADREVSEIS 528
QY 701 LMD-PNK-----FLLLVLQRYELAE 720
DB 529 IVDIANKRSSABESGQDVLENTFSQKHKLSVLLLEMKEAQBEIAFLKLQLOQKRAEEA 588
QY 721 FNKTISTKQDOLIKQ--NTLIEMLQVLIYIGERY----- 755
DB 589 DHEVL---DOKEMKQMEGEGIAPIKMKVLEDTGQDFPLMPNEESSLPVAVEKEQASTEHQ 645
QY 756 -----VPGVGNVTKEVMTREIHLICIEPMPHSA- 785
DB 646 SRTSEISINDAGVELKSTQDQDKSLSAVPDQCQHQDELE-RLKSOILELELNFHKAQ 704
QY 786 --IAKNLPENNETGLNVIKNKATFKPGVSGHGVY-----ELKDESL---KDFNMYF- 835
DB 786 --IAKNLPENNETGLNVIKNKATFKPGVSGHGVY-----ELKDESL---KDFNMYF- 835

DB 705 EYKENLDEKAKEISNLNOLIEE---FKNADNNSAFTALSEERDQLLSQVKELSMWTE 761
QY 836 --YHYSKTQHSKAEMQKRRKQENK---DEALPPPPPPFCPAFSKVINLLNCDIMMYI 890
DB 762 LRAOVKQLEMLNLAEEARQRLDYESQTAHDNLLTE-----QIHSLSTEAKSKDVKIEV 814
QY 891 LRTVFE-----RAIDTDSNLWTEGMLQMAFH 916
DB 815 LONELDDVQLOFSEOSTLIRSLQSQLOKNKESEVLEGAERVRHISKVEELSQAOLKE 874
QY 917 ILALG--LLEEK---QOLQKAPEE-----EVTFFYHKASRLG-----SSAMNIMLLEK 961
DB 875 ITKMDQLLEKRDVETLQOTIEKDQOVTEISFSTERKMWQLNEEKFSGLGVIEKTLKEQ 934
QY 962 LKGI-----PQLEGQKDMITWILQMPDTVKRLREKSLIVATTSSESINKNDIETHD 1013
DB 935 LNLSSRAEAKKEQVEEDNEVSSGLKNQYDEM-----SPAQGISKEELOHE 980
QY 1014 -----KKAERKRKAEEAARLHROKIMQAQMSALOKNFIEPHKLMYDNTSEMPGKEDSIMEE 1068
DB 981 FDLKKKENEORRKLQAALINKKELLQRYSLREE---ELANLKDESKKEIPLSETERGEV 1037
QY 1069 ESTPAVSQSYRIALGPKRGPSVTEKEVLTCLQEE-----BOEVKIENNAWVLS- 1117
DB 1038 EEDKENKEYSE-----KCVTSK-----CQEIYVLKOTISEKEVELQHIRKDLEE 1082
QY 1118 --ACVQKSTALTQHRGKPIELSGEALDPL 1144
DB 1083 KLAABEQQALVKQMNQTLQDKTNOIDLL 1111

RESULT 13

A57013
early endosome antigen 1 - human
N:Alternate names: endosome-associated protein
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A57013; S44243
R:Mu, F.T.; Callaghan, J.M.; Steele-Mortimer, O.; Stenmark, H.; Parton, R.G.; Campbell
J. Biol. Chem. 270, 13503-13511, 1995
A:Title: EEAL, an early endosome-associated protein. EEAL is a conserved alpha-helica
A:Reference number: A57013; MUID:95286647
A:Accession: A57013
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1410 <RES>
A:Cross-references: GB:I40157; NID:g1016367; PIDN:AAA79121.1; PID:g1016368
R:Seelig, H.P.
submitted to the EMBL Data Library, April 1994
A:Reference number: S44243
A:Accession: S44243
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-254; 'C', 256-257, 'LQ', 260-276, 'A', 278-283, 'A', 285-519, 'D', 521-574, 'EQ', 5
A:Cross-references: EMBL:X78998; NID:g475933; PIDN:CAA55632.1; PID:g475934
C:Genetics:
A:Gene: GDB:EEAL
A:Cross-references: GDB:1369996
C:Superfamily: human early endosome antigen 1
C:Keywords: calmodulin binding; endocytosis; metal binding; peripheral membrane prote

Query Match 2.0%; Score 186; DB 1; Length 1410;

Best Local Similarity 18.5%; Pred. No. 0.0033;

Matches 204; Conservative 200; Mismatches 424; Indels 272; Gaps 45;

QY 176 ENSRCPNEEVIVQARKIPPSVIKYVVMWTWEEKELPPELQIREKNERYCVLFDNDR 235
DB 275 ELSKGPQEVRYVQELQKLKSSVNLTKNQTLTNLLKKEQDYTKLEKH-----NEES 329
QY 236 HSYDHVITYSL-ORALDCELAQHLHTTAIDKEGRRAVKAQYAAQCAKEDIKSHSNVVS 294
DB 330 VSKKNIQATLHKDKDLCQQLQSLR-----SASE-TS 359


```
Db 815 -----LDTERSQAQERDAARQAQAQEQGKTALEQQAHEKVNQ-----L 858
QY 892 RTVFERADTDNSLWTEGMLQWAFHILALGLLEEQLOKAPKEEVEVTFDYHKASRLGSS 951
Db 859 REKWEK-----ERSWHOQELAKALESLEKEKMELEMLKEQOQTEMAIOAQAREERTQAE 913
QY 952 AMNIQMLEKLGIPQLEGQDMITWILQMTFTVRLKREKSLI--VATTSGSESIKDE 1009
Db 914 SALCOM-----QLETEKERSVLETLTQTELADASOOLERLQDMKVQKLKEQE 964
QY 1010 IT-----HDKAEKAEKRAEAEARHROKIMA---QMSALQKNFIETHKLMYDNTSEMPGKE 1062
Db 965 TTGILQTLQEAQRELK-EAARQHRDDLAALQOESESSILQDKMDLQKQVEDLKSQSLVAOD 1023
QY 1063 DS--IME---BESTPAVDYSRIALQPKRGVSTKEVUTCILCOEEQEVQKIENNAMWLS 1117
Db 1024 DSORLVEQVEKLETOEYNRIOKELER-----EKASLTLSLMEKEQRLVLQLEAD--S 1076
QY 1118 ACVOKSTALTOH---RGKPIELSGE-----ALDPLFMPDPLAYTYTGSCGHV 1162
Db 1077 IQQELSALRODMQBAQGEQKELSQMELLROEVKEADFLAQEAQLLEEL---EASHI 1133
QY 1163 ---MHAYCWQKYFAVQLSSOQRTHVDLFDLESGEYLCPLCKSLCNTVPIPILOPQKI 1218
Db 1134 TEQQLRASLWAQEAQAOL-----HLRLSTES-----QLEALAAEQ 1171
QY 1219 NSENADALQALLTLARWITQTVLARI 1243
Db 1172 PGNQAQAQALASLASALQAQALGSV 1196

RESULT 15
T26467
hypothetical protein Y11D7A.14 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T26467
R:Steward, C.
submitted to the EMBL Data Library, October 1998
A:Reference number: T20218
A:Accession: T26467
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1413 <WTL>
A:Cross-references: EMBL:AL032632; PIDN:CAA21588.1; GSPDB:GN00022; CESP:Y11D7A.14
A:Experimental source: clone Y11D7A
C:Genetics:
A:Gene: CESP:Y11D7A.14
A:Map position: 4
A:Introns: 17/3; 62/3; 122/2; 178/3; 200/1; 312/1; 351/3; 387/1; 463/3; 561/1; 583/3; 62
A:Superfamily: myosin heavy chain; myosin motor domain homology
F:35-696/Domain: myosin motor domain homology <NMO>
```

```
Query Match 2.0%; Score 180.5; DB 2; Length 1413;
Best Local Similarity 20.4%; Pred. No. 0.0073;
Matches 187; Conservative 127; Mismatches 350; Indels 253; Gaps 38;

QY 286 IKSHSENVSOHPLHVEVLHSEIHAQKALRLGSMNKMIM--SYSSDFRQIFQOACLREE 343
Db 336 IKVHDKLRKNNKONLAKTLSSASAMAKILYERLFGVIRKCNDAFVSD----- 382
QY 344 PDSENPCLISRLMDAKILYKARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLQKEY 403
Db 383 -DTESTCLRSRFI---AVLDIAGFEIIEKNSFEQFCINYNEKL--QQFNHMFQAEQ- 435
QY 404 ISDDHRSITISIALSVQMTVPTLARHLIEE-ONVISVITE-----TLLEVLPEY 452
Db 436 -SDYLEEGIKWTQVNFANHLQPTI--DLTEKPMGILSFLFEEECVPGNGSEKSLLEKLCN 492
QY 453 LDRNNKF-----NFQGYSDKLGRIYAVICDLKYLISK 486
Db 493 LNSDSFKKSKOTQKCTSTIRHFYVQHYAGEVHYNDGWLDKDRDNVETSVLD----ILSQ 548
```

Search completed: February 28, 2002, 10:02:40
Job time: 277 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 09:57:03 ; Search time 19.45 seconds
(without alignments)
2023.561 Million cell updates/sec

Title: US-09-724-126A-2
Perfect score: 9224
Sequence: 1 MADEEAGGTERMEISALFPQ.....ETARQETNOMLGFNQLL 1749

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	156.5	1.7	2710	1 US-08-480-604A-6	Sequence 6, Appli
2	156.5	1.7	2710	2 US-08-405-496A-6	Sequence 6, Appli
3	156.5	1.7	2710	4 US-08-915-136-6	Sequence 6, Appli
4	142.5	1.5	3248	1 US-08-353-700-1	Sequence 1, Appli
5	142.5	1.5	3248	5 PCT-US95-16216-1	Sequence 1, Appli
6	140.5	1.5	1708	1 US-08-493-092-2	Sequence 2, Appli
7	140.5	1.5	1708	1 US-08-508-836A-2	Sequence 2, Appli
8	139.5	1.5	2482	1 US-08-328-254-6	Sequence 6, Appli
9	139.5	1.5	976	4 US-09-104-324B-4	Sequence 4, Appli
10	133.5	1.4	3056	1 US-08-508-836A-8	Sequence 8, Appli
11	133.5	1.4	3056	2 US-08-629-001A-3	Sequence 3, Appli
12	133.5	1.4	3056	2 US-08-874-266-2	Sequence 2, Appli
13	133.5	1.4	3056	4 US-08-642-274D-3	Sequence 3, Appli
14	133.5	1.4	3056	4 US-08-952-127-3	Sequence 3, Appli
15	133.5	1.4	3056	4 US-08-952-014C-3	Sequence 3, Appli
16	132.5	1.4	1276	1 US-08-717-515-8	Sequence 8, Appli
17	131	1.4	1073	4 US-09-541-782-6	Sequence 6, Appli
18	130	1.4	2368	1 US-08-198-446B-15	Sequence 15, Appli
19	130	1.4	2368	2 US-08-870-693-15	Sequence 15, Appli
20	129.5	1.4	712	2 US-08-468-576B-17	Sequence 17, Appli
21	129.5	1.4	712	3 US-08-468-577B-17	Sequence 17, Appli
22	129.5	1.4	712	3 US-08-468-577B-17	Sequence 17, Appli
23	126	1.4	3672	2 US-08-822-445-12	Sequence 12, Appli
24	126	1.4	3801	2 US-08-822-445-10	Sequence 10, Appli
25	125.5	1.4	2154	2 US-08-841-349-4	Sequence 4, Appli
26	124	1.3	2860	2 US-08-826-267-2	Sequence 2, Appli
27	123.5	1.3	3959	2 US-08-970-269A-30	Sequence 30, Appli

28	123.5	1.3	3959	4 US-09-407-562-30	Sequence 30, Appli
29	123	1.3	876	1 US-08-717-515-4	Sequence 4, Appli
30	123	1.3	988	3 US-08-851-843A-69	Sequence 69, Appli
31	123	1.3	988	4 US-08-974-549A-112	Sequence 112, Appli
32	123	1.3	988	4 US-08-854-050-69	Sequence 69, Appli
33	123	1.3	1354	3 US-08-885-871-2	Sequence 2, Appli
34	123	1.3	2101	1 US-08-466-390-4	Sequence 4, Appli
35	123	1.3	2101	1 US-08-470-950-4	Sequence 4, Appli
36	123	1.3	2101	1 US-08-467-781-4	Sequence 4, Appli
37	123	1.3	2101	1 US-08-195-487-4	Sequence 4, Appli
38	123	1.3	2101	2 US-08-483-924-4	Sequence 4, Appli
39	123	1.3	2101	4 US-09-452-294-1	Sequence 1, Appli
40	123	1.3	2101	5 PCT-US93-06160-4	Sequence 4, Appli
41	122.5	1.3	1057	4 US-09-541-782-10	Sequence 10, Appli
42	121	1.3	577	4 US-09-315-794-32	Sequence 32, Appli
43	121	1.3	577	4 US-09-389-341-32	Sequence 32, Appli
44	121	1.3	816	2 US-08-533-306A-6	Sequence 6, Appli
45	121	1.3	816	2 US-08-742-923A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```

; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

Query Match 1.7%; Score 156.5; DB 1; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00041;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

Qy 283 KEDIKSHENVSQHPHVEV---LHSEIMAHOKFAIRLGSWMNKIMSYSSDFRQIFCOAC 339
Db 355 KSEIFSKLENLNVSDLEIKIAFALGVI---NQALISKQGSYLNLVTEQVKNRYQFLNQH 412

Qy 340 LREEPOSENCLISRLMLDAKLYGKARKILHELIFSSFFEMEMEYKKLFAMEFVKKYKQL 399
Db 413 LNPAESDNN-----FTDTTKIFHDSLFNSATAE---NSMFLTKIAPY---L 453

Qy 400 QKEYISDDHDSISITA-----LSVQMTFV-PTL-AHLIE---EQNVISVITE 443
Db 454 QVGFMPEARS-TISLSPGAYASAYDFINLQENTIEKTLKASDLIEFKPPENNLSQLTE 512

Qy 444 TLLEVLPEYLDNRKNFQCYSDKLGRYVA-----VICDLKYLISK-PTIWE 492
Db 513 QEINLSWSPQASAKYQFXYRYDITGGSUSGDNVFNKNTALDKNLLNNKIPSNVNE 572

Qy 493 RLRMQFLEGRSFLKILTCMQGMEIRRQVGHIEVDPDWEAATAIQMLKNILLMFQEW 552
Db 573 E-----AGSKNVHYHIIQGGDDISYEATCNLFKNP---KNSIIIOBNNESAKSY--F 622

Qy 553 CADDELLLVAYKECHKAVMRCSTSFSSSKTVVQSCGH-----SLETSYR 599
Db 623 LSDDGESIL---ELNK--YRIPELKNKEKVVTFIGHKDEFNTSEFARLSVDSLSNE 676

Qy 600 VSEDVSIHLPLSRTLAGLHVRLSRLGAVSRHLHEFFSDFQVEVLYEYPLRCLVLVAQV 659
Db 677 ISSFLDTIKLDISPK-----NVEVNLGCG-----NMFESY-DFNVE--ETYPGKLLLSIMDK 724

Qy 660 VA-----EMWRRNGLSLISOVFFYQDVKCREMY--- 688
Db 725 ITSTLPDVNKNSTIGANQYEVINSEGRKELLAHSGKWNKKEAIMSDLSKSEYIFDS 784

Qy 689 -----DKDIIIMQIGASLMDPNKFLLLVLQ-----RYELAE 719
Db 785 IDNKLAKSKNIPGLASISEDIKTLILDASVSPDKFILNNKLNIESIGDIYIYEKLE 844

Qy 720 AFNKTISTKQDLIKQYNTLIEMLQVLIYIGERYVPGVGNVTKKEVTHREIHLICIE 779
Db 845 PVKNIHNSIDDLIDFENLL-----ENVSDELYELKKL----- 877

Qy 780 PMPHSAIKNLPE-----NENNETGLENVINKVATFKKPCVSGHGVYELKDESLKD 830
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVIVETEKEI-- 919

Qy 831 FNMVFFHYSK-----TQHSKAEHMOKRRKQENKDEALPPPPPEFCFAPAFSKV 878
Db 920 FSKYSEHTKEISTIKNSIITDVNGNLLDNIQDHTSQ----- 957

Qy 879 INLLNCIDIMMYILRTVFERAIDTD--SNLWTEGMLQMAFHILALGLLEEKQOQLQAPEE 936
Db 958 VNTLNA---AFFIOSLIDYSSNNDVLDLSTSVKQVLAQLFSTGLNTIYDSIQLV--- 1010

Qy 937 VTDFETHKASRLGSSAMN---IQMLLEKLGKIPOLEGOKDMITW---ILOMFDTVKRLREK 991
Db 1011 -----NLSISNAVNNDITVLPITTEGIPTVSTILDGINLGAATKELLDHDPLLKK 1060
```

```

Qy 992 -----SCLIVATTSGSESISIKNDEITHDKAEARKKKAERKKAARLHROKIMQMS 1038
Db 1061 ELEAKVGVLAINNLSIAATVASIVGI-CAEVT-----IFLLPIAGISAGIP 1106

Qy 1039 ALOKNFIETHKLMYDNTSEMPCKEDSIMEEESTPAVSDYSRIALGPKRGSPVTEKE-VLT 1097
Db 1107 SLVNN-----ELIHDKATSVVYFNHLSSESKYGLPKTKEDDKILV 1147

Qy 1098 CILQOEQEVKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDILA 1151
Db 1148 PIDDLVISEIDFNNSIKLGTC---NILAMEGSGHTVTGN-IDHFFSSPSIS 1196

RESULT 2
US-08-405-496A-6
; Sequence 6, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-6
```

Query Match 1.7%; Score 156.5; DB 2; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00041;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

Qy 283 KEDIKSHENVSQHPHVEV---LHSEIMAHOKFAIRLGSWMNKIMSYSSDFRQIFCOAC 339
Db 355 KSEIFSKLENLNVSDLEIKIAFALGVI---NQALISKQGSYLNLVTEQVKNRYQFLNQH 412

QY	340	LREEPDSENPCLISRLMLWDAKLYGARKILHELIFSSFFMEMEYKKLFAFEMFYVYKQL	399
Db	413	LNPAISEDNN-----FTDTTKIHFSDLFNSATAE---NSMFLTKIAPY--L	453
QY	400	KQEYISDDHDRSISITA-----LSVMQFV-PTL-ARHLIE---BONVISVITE	443
Db	454	OVGFMPEARS-TISLSGPCAYASAYDFINQENTIENTXLTAKASDLIEKFPPENNLQSOLTE	512
QY	444	TLEVLPEYLDNRNKNFGQYSQDKLRVYA-----VICDLKYILISK-PTIWTE	492
Db	513	QEINSLWSFDQASAKYQFEKYVRDYTGGSLEDNGVDNFKNKTALDNKYLLNNKIPSNVE	572
QY	493	RLMQFEGFRFLKILCTCMQMGEIRIQVOGHIEVPDWEEAATAIQMLKANILLMFQEW	552
Db	573	E-----AGSKNYVHYIILOQGGDISYEATCNLFESKNP--KNSIIORNMESAKSY--F	622
QY	553	CACDEELLAVYKECHKAVMRCSSTFISSSTVVOSCGH-----SLETKSyr	599
Db	623	LSODGESIL-----ELNK--YRIPERLKNKEKVYTFIGHGKDEFNTSEFARLSVDSLSUNE	676
QY	600	VSEDLVSIHPLSRTLAGHLVRLSGAVSRLHEFVSPEDPQEVLEVYEURLCULVLVAQV	659
Db	677	ISSFLDTIKLIDSPK-----NVEVNLLGC-----NMFSY-DENVE--ETYPCKLILLSMDK	724
QY	660	VA-----EMWRNRGLUSLISQVYYODYVKCREMY-----	688
Db	725	ITSTLPDVNKNKSITIGANQOYEVRINSEGRKELLAHSKGWKINKEAINSDLSSKEYIFFDS	784
QY	689	-----DKDIIMLOICASLMDPNKFLILLVLO-----RYELAE	719
Db	785	IDNKLKAKSKNTPGLASISEDIKTLTLLDASVPDTKFLNNLKLINESSIGDYIYYEKLE	844
QY	720	AFNKTISTDQDLIKOYNLTILEMLQVLIYIVGERYPGVGNVTKEETWREIHILLCIE	779
Db	845	PVKNIHNSIDLDIEFNLL-----ENVSELEYELKKL-----	877
QY	780	PMPSIAIAKNLPE-----NENNETGLENVINKVATFKKPGVSGHGVIYELKDESJKD	830
Db	878	-----NNLDEKYLISFEDISKNNSTYSYRFINK-----SNGESVYVETEKEI--	919
QY	831	FNMVFYHYSK-----TQISHKAHQMKRRKOENKDEALPPPFPFCPAFSKV	878
Db	920	FSKYSHITKEISTIKNSIIDTVNGNLNDINTLOQHTSO-----	957
QY	879	INLLNCDIMMYTLRTVFERAITDO--SNLWTEGMLOMAFHIALGLLEEKOQOKAPEEE	936
Db	958	VNTLNA--AFFIOQLDIYSSKNVDNLDSITSVKVOLYAQLFTGLNTIYDSIQLV----	1010
QY	937	VTDFDYHKASRLGSSAMN--IQMLEKLKGPOLLEGQKDMITW---ILOMFDTVKRLIREK	991
Db	1011	-----NLISNAVNDTINVLPITTEGPIPVSTILDGINLGAAIKELLDEHDPLKK	1060
QY	992	-----SCLIVATTSGSESINKDEITHDKKAERKKAEAARLHRQKIMAQMS	1038
Db	1061	ELEAKYGVLAINMSLSIAATVASIVGI-GAEVT-----IFLLPIAGIASAGIP	1106
QY	1039	ALOKNEIETHKLWYDNTSEMPCKEDSIMBEESTPAVSYSRIALGPKRGSPSVTEKE-VLT	1097
Db	1107	SLVNN-----ELIHLDRKATSVVNTFNLHSESKYGPGLKTEDDKILY	1147
QY	1098	CILCOBEQEVKIENNAMVLSACVOKSTALTQHGRKPTELSELGEALDPLFMPDILA	1151
Db	1148	PIDDLVISEIDFNNSIKLTCG----Ntlameggsghtvtgn-IDHFESSPSIS	1196
RESULT 3			
US-08-915-136-6			
; Sequence 6, Application US/08915136			
; Patent No. 6250960			
; GENERAL INFORMATION:			
; APPLICANT: KINK, JOHN A.			
; APPLICANT: THALLEY, BRUCE S.			

```

> APPLICANT: PADHYE, NISHA V.
> APPLICANT: FIRCA, JOSEPH R.
> TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
> PREVENTION OF C. DIFFICILE DISEASE
> NUMBER OF SEQUENCES: 32
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: MEDLEN & CARROLL, LLP
> STREET: 220 MONTGOMERY STREET, SUITE 2200
> CITY: SAN FRANCISCO
> STATE: CALIFORNIA
> COUNTRY: UNITED STATES OF AMERICA
> ZIP: 94104
>
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.30
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/915,136
> FILING DATE:
> CLASSIFICATION:
> PRIOR APPLICATION DATA: 08/480,604
> FILING DATE:
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/405,496
> FILING DATE: 16-MAR-1995
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/329,154
> FILING DATE: 25-OCT-1994
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/161,907
> FILING DATE: 02-DEC-1993
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 07/985,321
> FILING DATE: 04-DEC-1992
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 07/429,791
> FILING DATE: 31-OCT-1989
> ATTORNEY/AGENT INFORMATION:
> NAME: INGOLIA, DIANE E.
> REGISTRATION NUMBER: 40,027
> REFERENCE/DOCKET NUMBER: OPHD-01763
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (415) 705-8410
> TELEFAX: (415) 397-8338
> INFORMATION FOR SEQ ID NO: 6:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 2710 amino acids
> TYPE: amino acid
> TOPOLOGY: linear
> MOLECULE TYPE: protein
>
> US-08-915-136-6
>
> Query Match 1.7% Score 156.5 DB 4; Length 271
> Best Local Similarity 18.6%; Pred. No. 0.00041;
> Matches 189; Conservative 154; Mismatches 354; Indels 31
>
Qy 283 KEDIKSHSNVSOHPHLHVEV---LHSEIMAHOKFALRLGSGWNKIMSYSSDFRIG
Db 355 KSEIFSKLENLVNSDLEIKTAFALGSGVI--NQALISKQGSYLTNLVIEQVKRNYQ
Qy 340 LREEPDSENPCLISRLMLWDAKLYKGARKILHELIFSFSEFMMEYKFLFAMEFVK
Db 413 LNPATESDNN-----PTDTTKIFHDSLFNSATAE---NSMFLTKIAP
Qy 400 KREYISDDHRSISITA-----LSVQMFTV-PTL-ARHLLE---EQNVII
Db 454 QVGFMEPKRS-TLSLSPGAYASAYDFINQENTIEKTLKASDLTLEFKFPENNLL
Qy 444 TLLEVLPEYLDNRNKNFNGOYSQDLKGRVYA-----VICDLKYIILSK-P

```

Query Match	1.7%	Score 156.5;	DB 4;	Length 2710;
Best Local Similarity	18.6%;	Pred. No. 0.00041;		
Matches 189;	Conservative 154;	Mismatches 354;	Indels 317;	Gaps
Qy	283	KEDIKSHENVSQHLHVEV---LHSEIMAHQKFAFLGSMWNKIMSYSSDFRQIFCQAC	339	
Db	355	KSEIFKLENLVSDLEIKAFALGSI---NQALISQGSYILNLVIEQVKNYQFLNQH	412	
Qy	340	LREEPDENPCLISRLMLDAKLYKGARKITHELIESFFMEYKKLFAMEVVKYKQL	399	
Db	413	LNPAIESDNN-----FTDPTKTFHDSLFSNATAE---NSMPLTKIAPY---L	453	
Qy	400	QKEYISDDHRSISITA-----LSVQMTFV-PTL-ARHLIE---EQNVISVIYE	443	
Db	454	QVGMPEARS--TISLSPGAYASAYDYDFINLQENTIEKTLKASDLIEFKPEPNLSQLTE	512	
Qy	444	TLLEVLPEYLDNRNKNFQGYSDOKLGRVYA-----VICDLKYIILSK-PTIWI	492	

RESULT 3
US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.

Db 513 QINLSWSPQASAKYQFYKDYVTGGSLSDNGVDKFNKNTALDKNYLLNKNKIPSNVVE 572
QY 493 RLRMFQLEGRSFKILTLCMQMEEIRROVQGHIEVDPOWEAAIATOMOLKNILLMFQEW 552
Db 573 E-----AGSKNVVHYIOLGDDISYEATCNLFKSNP--KNSIIQRNMNESAKSY--F 622
QY 553 CACDELLLVAYKECHKAVNRGCTSTFSSKTVVQSCGH-----SLETKSYR 599
Db 623 LSDGESIL-----ELNK--YRIPERLKNKEKVVTIGHKDEFNTSEPARLSVDSLSNE 676
QY 600 VSEDVSLHPLSLRTLAGLHVLRLSRLGAVSRHLEFVSFEDQVEVLVEYPLRCLVLVAQV 659
Db 677 ISSFLDTKLDISPK-----NVEVNLGC-----NMFYSY--DFNVE--ETYPGKLLLSIMDK 724
QY 660 VA-----EMVRRNGSLISQVYFYQDVCKREEMV--- 688
Db 725 ITSTLPDVNKNITIGANQYEVIRNSEGRKELLASHGKWKINKEEATMSDLSKEYIFFDS 764
QY 689 -----DKDIIMLOIGASLMDPNKFLLLVLO-----RYELAE 719
Db 785 IDNKLKAKSNIPGLASISDEIKTLLDASVSPDTKFLNKLKNTFESSIGDYIYYEKL 844
QY 720 ANKTIKQDOLIKOYNLTIEMLQVLIYVGERVYVPGVGNVTKKEVTMRIIHLCLIE 779
Db 845 PVKNIHNSIDDLIDFNLL-----ENVSDLEYELKKL----- 877
QY 780 PMPHSAIAKNLPE-----NENNETGLENVINKVATFKPGVSGHGVYELKDESLKD 830
Db 878 -----NNLDEKYLISFEDISKNNSTSVRFINK-----SNGESVYVETEKEI-- 919
QY 831 FMNYTHYSK-----TQHSKABHMQKRRKQENKDEALPPPPPPFCPAFSKV 878
Db 920 FSKYSEHITKEITKNSIITDVNGNLLDNIQLDHTSQ----- 957
QY 879 INLLCNDIMYILRVFERAIDTD--SNLWTEGMLQMAFHIALGLLEKQLOKAPBEE 936
Db 958 VNTLNA-----AFFIQSLIDYSSKNDVNDLSTSVKQVLAQLFSTGLNTIYDSIQLV-- 1010
QY 937 VTFDFVHRASRIGSSAMN--IOMLEKLGKIPOLEGQKDMITW---ILOMFDTVKRLREK 991
Db 1011 -----NLISNAVNTINVLPTITGIPVISTILDINGINLGAIAKELLDEHDP LKK 1060
QY 992 -----SCLIVATTSGESIKNDEITHDKEAERKKAEPAAHLRQKIMQAOMS 1038
Db 1061 ELEAKGVLAINNLSIAATVASIVGI--CAEVT-----IFLLPIAGISAGIP 1106
QY 1039 ALOKNIETHKLYMDNTSEMPGKEDSIMPEESTPAVSYSRIALGPKRGPSTVEKE--VLT 1097
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSEKKGYP LKTEDDKILV 1147
QY 1098 QTLCOEEOBKVIENNAMVLSACVOKSTALTQHRGKPIELSGEALDPLFMDPDLA 1151
Db 1148 PIDDLVISEIDFNNSIKLGTG-----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196

RESULT 4

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA

; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353.700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 1.5%; Score 142.5; DB 1; Length 3248;
Best Local Similarity 18.2%; Pred. No. 0.01;
Matches 302; Conservative 245; Mismatches 601; Indels 515; Gaps 76;

QY 166 HEPGRAGTIFKNSRCPLNEE---VIVQARKTFPSVIKVVEMTWEEB---KELPPELQ 218
Db 304 HEREMKGOVNFQELQLEKAKVELIEKEKVLNKCDELVTRTTAQYDQASTKYTALEOK 363
QY 219 IREKNERYVYVLFNDE--HHSYDVHYISLQALDCELAQAQLHTTAIDKEGRRVAKAGAY 276
Db 364 LKLTEDLSQORNAESARCSLEQKIKEKEFEQELSROORSFOTLQECIQ--MKARLT 422
QY 277 AACQEAEDTKSHSENSVQHPHVEVLHSEIMAHOKFALRLGSGWNKIMSYSSDFRQIFC 336
Db 423 QELQQAQN-----MH--NVLOAELD-----KLTSVKQOENLNLEBEFKQLC 461
QY 337 QACLREEPDSENPCILSRLMLWDAKLYKGARKILHELIFSPFEMEMKYLKFAMEFVKY 396
Db 462 RA-----EQAFOAIOIKENELRR---SMEEMKKE 487
QY 397 KQLQKEYISDDHRSISITALSVMFTVP---TLARHLIEQONVISVITETLLEVLPEYL 453
Db 488 NLLK---SHSEQAKAREVCHLEAEKLNKQCLNOSONFAEEMKAKNTSOETMLRDLQE-- 542
QY 454 DRNNKFNQGYSDKLGKRVYAVICDLKYLISKPTIWTIRLMOFLEGFRSLKILTQMQ 513
Db 543 ----KINQOENSL-TLEKLLKLVADLE-----KQDCSQ 571
QY 514 GMEIEIRQVGOHIE-----VDPDWEAAI-AIQM-----OLKNILLMFQEWACADEE 558
Db 572 DLLKRE---HHIEQLNDKLSKTEKESKALLSALAEKKEEELKEETLFSWSENEK 628
QY 559 LLLVAYKECHKAVNRGCTSTFSSKTVVQSCGHSLSTKSYRVSDELVSITHLP---LSRTL 615
Db 629 LL-----TQMESEKENLQSKINHLCTKQTKQIKSHYENRVRTLEMDRENLSVEI 679
QY 616 AGHL-----VRLSRLGAVSRHLEFVSFEDQVEVLVE--YPLRCLVLVAQVVAEMWR 665
Db 680 RNLHNVLDSSKSVETQKL--AYMELQKAEFSQDKHQKEIENNMKLTQSQTQVED---L 735
QY 666 RNGLSLISQVYFYQDVCKREEMV-----KDIIMLOIGASLM---DPNKFLLLVQRYEL 717
Db 736 EHKLQLLSLIMDKD--RCYQDLHAIEYSLRDLKSK--DASLVTNEDHORSLLAFDQOPAM 793

QY 718 AEFANKTI-----STKDODLI-----KOYNTLIEML 744
Db 794 HHSFANIIEGSGMPSECRLEADQSPKNSAILQNRVDSLEFSLESOKMNSDLQOC 853
QY 745 QVLIYIVGVRYPGVNVTKEVTMRILHLLCI-----EWPMP-SATAKNLP 791
Db 854 EELVOIKGE-----IENLAKMAEQMHOSFVAETSORISKLOEDTSAHONVVAETLS 904
QY 792 ENNETGLENVINKVAT-----FKPGVSGHGVYELKDESLKDFNNFYHYSKTQHS 844
Db 905 ALENKEKELQLLNKVKETEQAIEOLKK---SNH-----LLEDSLAKELQLL-----S 948
QY 845 KAEHMQKRRKQENKDEALPPPPPPFCFAPFSKVINLLNCNDIMMYILRTVFERAID--TD 902
Db 949 EFTLSLEKKE-----MSSIISL-----NKREIEELTQ 974
QY 903 SNLWTEGMLQAFILALGALLEEKQOQKAPBEVTFDYHAKASRLGSSAMNIQMLEKL 962
Db 975 EN---GTLKE-----INASLNQERKNMLQKSE---SFANYIDEREKSTISELSQYKQEKL 1023
QY 963 KGIPOLEGQKDWITWLOMFTVRLREK-SCLIVATTSGSESINKD---EITHDKKA 1017
Db 1024 ILLQCEETGNAYEDLSOKYKAAQKNSKLECLNNECTSLCENRKNLEQLKEAFKEHQ 1083
QY 1018 ERKRAEAARLHROKIMQMSALQKNFTETHKLMVDNTSEMPGKEDSIMEESTPAVSQY 1077
Db 1084 EELTKLAFABERNQNLMLETVQO---ALRSEMTDNQNNKSEAGLQKEIMT----- 1134
QY 1078 SRIALGPKRGPSVTEKEVLTICLOEBQEVKIENNAMVLSACVQKSTALTQHRGKPIELS 1137
Db 1135 -----LKEQNKMOKEV-NDLQENQMLKVMK-----TKHE----- 1165
QY 1138 GEALDPLFMDPLAYGTYTGCGHVMAVCQKYFEAVOLSSQRIHVLDLPLESEYLC 1197
Db 1166 -----CONLESEPIRNSVKER-----ESE----- 1184
QY 1198 PLKSLCNTVPIPILOPKINSENADA-LAQLLTAR-----WIQT 1238
Db 1185 ---RNOCN-FKQMDLEVKETISLDSYNAQLVQLEAMLNKELKLOESKEKECLOHEQOT 1240
QY 1239 VIARISGYNIRHAKENPDIPIFFNOGMDSTLEPHSILSFGVSESIKYSNIKEMVILFA 1298
Db 1241 IRGDLTSLNLDMSQSE-----ISGLKDCETDAEBEKYISGPHELSTSONDAHLCQCSIQ 1294
QY 1299 TIYIRIGLKVPPDERPRVPMLTWSTCAFTQATENLLGDECKPLFGALQNRQHNGLKAL 1358
Db 1295 TTMNKLN-----ELEKICEILOAEKYELVTELNDSRSECITAT 1332
QY 1359 MQFAVAORITCPQVLIQKHLVRLSVLPNIKSEDTPCLLSIDLPHVLVGAVALAPPSLYW 1418
Db 1333 RKMA-----EEVKLLNEV--KLNDDSG--LLHGELVEDIPGGEFG-EQPN 1375
QY 1419 DDPVDLPSSVSSYNHLYLPHLITMAHMLQILITVDTGLPLAQVQDESEAH-SASSFF 1477
Db 1376 QHPVSLAPLDESNSYHEHTLSDKEVQMHFAEL---OEKFLSLQSBKILHDOHCQMSKM 1432
QY 1478 AEISQYTSG-----SIGCDIPGWY-----LWVSLKNGITPYLRCAAL----- 1514
Db 1433 SELQTYVDSLKAENLVLTNURNFGDLVKEMQLGLEGLVPSLSSSCVPSSSSLSLGD 1492
QY 1515 ---FFHYLLGVTPPEELHTNSAEGVSA-----LCSYLS-----LPTNLF 1552
Db 1493 SSFYRALLEQTDGMSLLSN-LEGAVSANQCSVDEVFCSLSQTYVDSLKAENLVLTNLR-R 1550
QY 1553 LFQEWYDTR-----PLLRQWCA-DPALLNCLKQKNTVVYPRKRNLSLIELPDD 1600
Db 1551 NFQ--GDLVKEMQLGLEGLVPSLSSSCVPSSSSLSLGDSSFY-----RALLEQTDG 1601
QY 1601 YSCLLNQASHRCRPSRADDKRPVLCILFCGAILCSQNICQOE 1643
Db 1602 MSLLSNLGGVVSANQCSVDE-----VFCSS-LQEENLTRKE 1636

RESULT 5
PCT-US95-16216-1
; Sequence 1. Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 1.5%; Score 142.5; DB 5; Length 3248;
Best Local Similarity 18.2%; Pred. No. 0.01;
Matches 302; Conservative 245; Mismatches 601; Indels 515; Gaps 76;
QY 166 HEPGRAGTTKNSRCPLNEB-----VIVQARKIPSPVVIKYYVEMTWEEB---KELPPELQ 218
Db 304 HEKEMGQVKNKFOFLOLQLEKAKVELKEKVLNKRDELVRTTAQYDQASTKYTALEQK 363
QY 219 IREKNERYCVLPNDE--HHSYDHVYISLQALDCELAELQHTTATDKEGRRRAVKAGAY 276
Db 364 LKLTEDLSQORQNAESARCSLEQKKEKEFEQELSRQORSFQTLDDQECIO-MKARLT 422
QY 277 AACQAEKEDIKSHSNVSOHPLHVEVLHSHMAHQFALRLGSMNKMYSYSSDFQITFC 336
Db 423 QELQQAQN-----MH-NVLAELD-----KLTSVKQOLENNLEEFKQKLC 461
QY 337 QACLREEPSNENCLISRLMLWDAKLYKGARKILHELIFSFFMEMEYKLLFAMEFVYKY 396
Db 462 RA-----EQAFQASQIKENELRR--SMEEMKKE 487
QY 397 KQLOKEYISDDHRSISITALSVQMTVP---TLARHLIEEQNVISVITETLLEVLPEYL 453
Db 488 NNLLK---SHSEOKAREVCHLEAELKNKQCLNQSOFABEMKAKNTSQETMLRDLOE-- 542
QY 454 DRNNKFNFGYSGDKLGRVYAVICDLKYILISKPTIWTWRLRMQFLGGRSFLKILCMQ 513

Db 543 -----KINQENSL-TLEKLKLVADLE-----KQDCSQ 571
QY 514 GMEIRROVQGHTE-----VDPWEAAI-AIOM-----OLKNILLMFOEWACADEE 558
Db 572 DLLKRE---HHTEQLNDKLSKTEKESKALLSALEKLKKEVEELKEKTTLFSCWKSSENEK 628
QY 559 LLLVAYKECHKAVMRCSTSFISSKTVQSCGHSLETKSVRSVEDLSIHLPL---LSRTL 615
Db 629 LL-----TOMSEKENLOKINHLETCLTKQIKSHENYRVRTLEMDRENLSVEI 679
QY 616 AGHL-----VRLSRGAVSRLEHFEVSFDFQVBLVE-YPLRCLVLVAOVAMWR 665
Db 680 RNLHNVLDKSEVEYQKL-AYNELOQKAFFSDOKHOKETENMCLTKTSQLTQGVED---L 735
QY 666 RNLGLISQVYFYQDYKCREMYD-----KDIIMLOIGASLM---DPNKFLLLVLQRYEL 717
Db 736 EHKQLLSNEIMDKD-RCYQDLHAESLRDLKSK-DASLVTNEDHORSLLAFDQOPAM 793
QY 718 AEFNKTI-----STKDQDLI-----KQYNTLFEML 744
Db 794 HHSFANIIGEQSGMPSESRLEADQSPKNSAILQNRVDSLEFSQKOMSDLQKQC 853
QY 745 QVLIYIVGRYVGVGNVTKEEVYTMREIHLICI-----EPMHP-SAIKPNP 791
Db 854 EELVQKGE-----TEENLMKAQOMHQSFVAETSQRIKLOEDTSAHQNVVAETLS 904
QY 792 ENENNETGLENVINKVAT-----FKKPGVSGHGVYELKDESLKDFNMFYHYSKTQHS 844
Db 905 ALENKEKELQLNDKVETEQAETQLKK---SNH-----LLEDSLKELQLL-----S 948
QY 845 KAEHMOKKRRKQENKDEALPPPPPECPAFKVINLLNCIDIMMYILRTVFERAID--TD 902
Db 949 ETLSLEKKE-----MSSIISL-----NKREIEBELTQ 974
QY 903 SNLWTEGMLQMAFHIALGILLEKQOOLAKAPEBEVTFDYHKASRLGSSAMNIQMLLEKL 962
Db 975 EN-----GTUKE-----INASUNQKMLIQKSE---SFANTIDREKSISELSQYQKEL 1023
QY 963 KGIPOLEGOKDMITWILQFDTVKRLREK-SCLIVATTSGSESIKND-----EITHDKEKA 1017
Db 1024 ILLQRCETGNAYEDLSQYKAAQERNKLECLNCTSLCENKRNKELEQLEKFAKEHQ 1083
QY 1018 ERKRKAARLHRQKIMAOXSALQKFIETHKLMYDNTSEMPKEDSIMEESTPAVSDY 1077
Db 1084 EFLTKLAFAEERNQNLMELETVQO---ALRSEMTDNQNNKSEAGGLKQEIIMT----- 1134
QY 1078 SRTALGPKRGPSVTEKEVLTCLICEQEOEVKIENNAWLSACVQKSTALQHRKPIELS 1137
Db 1135 -----LKEEQNKQKEV-NDLLOENQOLMKVMK-----TKHE----- 1165
QY 1138 GEALDPLFMDPDLAYGTYTGSCGHVMHVCWKYFEAVQLSSQOIRHVDLFDLESGEYLC 1197
Db 1166 -----CONLESEPIRNSVKER-----ESE----- 1184
QY 1198 PLCKSLCNTVPIPILOPOKINSINADA-LAQLLTLLAR-----WIQT 1238
Db 1185 ---RNQC-FKPMQDLEVEKISLDSYNAQLVQLEAMLRNKELKLOEKEKECQLHELOT 1240
QY 1239 VLARISGYNIRHAKGNPIPIFNQGMGSTLEFHSITLSPGVESSIKYSIKEMWILFA 1298
Db 1241 IRGDLETSNLQDMQOE-----ISGLKCEIDAEEKYISGPHELSTSQNDNAHQCSLQ 1294
QY 1299 TTYIRGLKVPDPDRPRVPMLTWSTCAFTIOAIENLLGDEGKPLFGALONRQHNGLKAL 1358
Db 1295 TTNKLN-----ELEKICEILOKEKVELVTELNDNRSEICITAT 1332
QY 1359 MOPAVAQRITCPQVLQKHLVRLLSVPLNPKSEDPTPCLLSIDLPHVLVGAVALFPFSLYW 1418
Db 1333 RKMA-----EEVGKLLNEV-KILNDDSG-LLHGLVEDIPGGEFG-EQPNE 1375
QY 1419 DDPVDLOPSSVSSSYNHLFIETHMAHMLQILLTVDTPGLPQAQVEDSEAH-SASSFF 1477
Db 1376 QHPVSLAPLDESNSYEHLTLSDKREVQMHAEL---QEKFLSLQSENKILHQHCOMSSKM 1432

QY 1478 AEISOYTSG-----SIGCDIPGWY-----LWVSLKKNITPYLRCAAL----- 1514
Db 1433 SELQTYVDSLKAENLVLTNLRNFQGLVKEMQLGLEGLVPSLSSCVPDSSSLSLGDD 1492
QY 1515 --FHYLLGVTPPELHTNSAGEYSA-----LCSVLS-----LPTNLFL 1552
Db 1493 SSFYRALLEQDGMSSLN-LEGAVSANOCSDVEFCFCSLQTYVDSLKAENLVLTNLR 1550
QY 1553 LFQEVWDTVR-----PLLQWCA-DPALLNCLKOKNTVVRYPRKRNLSLIELPD 1600
Db 1551 NFQ---GDLVKEMQLGLEGLVPSLSSCVPDSSSLSSLDGSSFY-----RALLEQTD 1601
QY 1601 YSCLLNQASHFCRPSADDERKHPVLCFLFCGAILCSQNICCOE 1643
Db 1602 MSLLSNLEGVVSVANOCSDVE-----VFCSS-LOEENLTRKE 1636

RESULT 6

US-08-493-092-2
; Sequence 2, Application US/08493092
; Patent No. 5728807
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Taglie, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-Telangiectasia Gene
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/493,092
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-310 (TAV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1708 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-493-092-2

Query Match 1.5%; Score 140.5; DB 1; Length 1708;
Best Local Similarity 17.9%; Pred. No. 0.0053;
Matches 260; Conservative 197; Mismatches 469; Indels 529; Gaps 68;

QY 319 SWMKINYSDDFRQFQOACLR-----EPDSENP-----CL---ISRLMLWD 359
Db 509 SWRNLSTHV---QGFETSLRHFQSRTSTPANLDBSEHFFRCCLDKKKSQRTMLAV 564
QY 360 AKLYKGARKILHELIFS-SPEMEMEYKKL-----FAMEFVKYKQLOKEVISDD 407
Db 565 VDYMRORRPSSTGTFIDAFWLDLNLVLEAKVAOSCAAHFTALLYAEIY--ADKSMDDQ 622
QY 408 HDRSISITALSVQMTPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNFQGSOD 467

Db 623 EKRLAFEGS-QSTTSSUSEKSEETGI--SLDLLLLLEIYRSIGEPDSLYGCGG----- 675
Qy 468 KLGWYAVICDLKYLISKPTIWERLMOFLEGRFLKILTCMOGMEETROVGO-HI 526
Db 676 --GKWLQPIELR--TYEHEAMWKAALVYDLE---TAIPSTQAGIIQALQNLGLCHI 728
Qy 527 EVDPDWEAIIAQMOLKNILLMFOWCACDELLVA-----YKECHKAYMRC-STSF 579
Db 729 -----LSVLKGLDYENKDWCPLEELHYQAARNQMDHCTSVSKEVEGTSYH 777
Qy 580 SSSKTVVQSCG-----HSLFKTSRVSDEDLVSIHLPLSRTLAGLHVRLSRLGAVR 631
Db 778 ESLYNALQSLRDREFTFYESLKYARVKEVEEMCKRSL--ESVYSLYPTLSRLQAI 834
Qy 632 HEFVSFEDFQV-----EVLVEYPLRCLVLAQVVAEMWRNGLSLISQVFFYQDV--- 681
Db 835 ESIGELFSRVTQRQLSEVIYK-----WQHSOLLKDSDFSQEPIMAL 878
Qy 682 -----KCREMYDKDIIMLQICASLMDPNKFLLLVLQRYELAEAFNK 723
Db 879 RTVILLEILMEKEMDSQRECIKDILTKHLVELSI-----LARTFKN 919
Qy 724 T-ISTKODLIKQNTL-----IEMLOVLIYIGERYVPGVGNVTKKEVTMREI 774
Db 920 TQLPERAIFQIKQNSVSCGVSEWQLEE-AQVF-----WAKKEQSLASLIL 964
Qy 775 LLCTEPHPSAIKN-----LPENENNETGLENNVINKVATPKPKGVSGHGVY- 821
Db 965 KOMIKKLDASCAANNPSLKLYTTCRLVCGNWLAECLLENPAVIMQTYLXAVEVAGNYD 1024
Qy 822 -----ELKDESLKDFNMYFYHYSTQHSKAEMHOK-----KRRKOE----- 857
Db 1025 GESSDELNRGKMAF-LSLARFSDQYQRIENYMKSEFENKQALLKRAKEEVGLLREHK 1083
Qy 858 -----NKOE-----ALPPPPPPFECFPAFSKVINLLNCDIM-----MYILRTV 894
Db 1084 IQTNRYTVKVORELEDELALRALKEDKRPCKA---VENYINCLLSGEEHDMWVRL- 1139
Qy 895 FERAIDTNSLWTEGMLQMAFHIALGLLEKQOLKAPPEEVTFDP-----YHKASRLGS 950
Db 1140 -----CSLWLENS-----GVSEVNGMKRDMGKIPITYKFLPLMYQLAARMGT 1181
Qy 951 SAMQIOMLEKLGIPOLEGOKMITWLOMFDVTKRLREKSLIVATVTSSESINKDEI 1010
Db 1182 KWMG-----GLGPEVLNLLISRI-----SMDHPHHTLFIILANAN-----RDEF 1223
Qy 1011 THDKEKAERK-----KAEAA-----RLHRQKIMAQMSALQKNFI--- 1045
Db 1224 LTKPEVARRSRITKNVPKQSSQLDRETEAANRICTIRSRPQWRSVEALCDAYIILA 1283
Qy 1046 -----ETHKLMYDNTSEMPCKEDSITWEEESTPAV-----SDYS-----R 1079
Db 1284 NLDAQMTQKRGKINIFADQPIITLKNLEDVYVPTMEIKVDHTGEYNLVTIQSKAEFR 1343
Qy 1080 IALGPKRGPSVTEKEVLTCI-----LCOEEOEVKLEN 1111
Db 1344 LAGG-----VNLPKIIDCVSGDKERRQLVGRDDLQDAVMQVQFOMQMTLLQRTET 1397
Qy 1112 NAMVLSACVOKSTALTQHRG-----KPIELSGEALDPLFMDPDLAYGYTGTSCGHVM 1163
Db 1398 RKRLTICTYKVPVLSQSGVLEWCTGTVPPI---GEF---LVNEDGAHKRYRPNDFSAP 1451
Qy 1164 HAVCWQKYFAVQLSSQORIHVDLFDLESGEYLCPLCKSLCNTVPIPIPOQKINSENA 1223
Db 1452 Q--CQKKMVEYQKKSFEKEVEFM-----DVCQNEQPVF-----RY 1485
Qy 1224 DALAQLLTLARWIQTLA-----RISGVNIRHAKGENPIPIFFNOGMDSTLEPHSI 1275
Db 1486 FCMKFKLDPALWFKRLAYRTSVATSSIVGYL-----GLGDRHVQ----- 1526
Qy 1276 LSFQVSESIKYSNSIKEMVILFATTIYRIGLKVPPDERDP-----RVPMILT 1321

Db 1527 -----NLLINEQSAELVHIDLGVAFEQCKILPTPETVPFRULTRDIVDGMGITVEGVF 1579
Qy 1322 WSTCAFTTQAIEN-----LLGDEGKPLFGALQNRHNGLKALMQFAVAQRITCP 1370
Db 1580 RCCCKTHEVMNSQETLLTIVEVLLYD---PLF-----DWTMNPALKALY---LQOR---P 1626
Qy 1371 QVLIQKHVLRLLSVVLPNIKSEDTPCLLSIDLFHVLVGAFLAPPSLYWDDPVDLPQSSVS 1430
Db 1627 EDETELH-----PTLNADDQECKRNL-----SDID 1651
Qy 1431 SSYNHLYFLHLLTMAHMLQIILLTVDTGLPLA---QVQ---EDSEEAHSASSFAEISQYT 1484
Db 1652 QSFQDVAERVLMLRQLEKLR---GVEEGTVLSVGGQVNLIIQQAIDPKNLSRLF----- 1701
Qy 1485 SSGICGDIPGWYLVW 1499
Db 1702 -----PGWKAWV 1708

RESULT 7
US-08-508-836A-2
; Sequence 2, Application US/08508836A
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-Telangiectasia Gene
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,836A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-313 (TAU)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1708 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-508-836A-2

Query Match 1.5%; Score 140.5; DB 1; Length 1708;
Best Local Similarity 17.9%; Pred. No. 0.0053;
Matches 260; Conservative 197; Mismatches 469; Indels 529; Gaps 68;

Qy 319 SWNNKIMYSDFRQIFCQACLR-----BPDSNP-----CL---ISRMLWD 359
Db 509 SWRNLSTHV---QGFTTSLRHFQTSRSTTPANLDSESEHFFRCCLDKKSORTMLAV 564
Qy 360 AKLYKGARKILHELIFS-SFFMEMEYKKL-----FAMEFYKYLQKQKEYISDD 407
Db 565 VDMRRQRKRPSSGTFINDAFWLDLNLVAKVAQSCAAHFTALLYAEIY--ADKKSMDQ 622

QY 600 VSEDL---VSIHLPLSRTLAGLHVRLSRGAVSRUHEFVSFDFQV-----EV 644
Db 306 QSENKQSTKQSHLTKELEDKVSQSVSTQKALE-----EDLQATATKICQLTEKET 361
QY 645 LVEYPLRCLVLVAOVVAEMRRNGLSLSQVYVQDVKCR-EMYDKDILMIQIGASLMD 703
Db 362 QMEENKARAHSFVTE-----FETVCSLELLTEQORLEKNE---D 403
QY 704 PNKFLLLVLQR--YELAEAFNKTIKTDODLIKQVNTLIEEMQLVLIYVGRYVPGVN 761
Db 404 QLKILTMELQKSSSEL-BEMTKLTNNKEVEL-----EELKKVL---GEKETILYEN 450
QY 762 VTKEEV-----TMBRIIHLICIEPMPHSAIAKNLPENNETGL---ENVINKVATFKK 812
Db 451 KOFETIAELKTEQELIGLL-----QAREKEVHDLEIQLTAITTSQYYSKEVKDLK 503
QY 813 PGVSGHGYVELKDESJKDFNMFYHYHSKTOHSAEHMOKRRKOEKDEALPPPPPERC 872
Db 504 T-----ELENKLNTEL-----TSHCNKUSLENKELTQETSDMTLELKNQOEDI 548
QY 873 -----PAFSKVINLLNCDIMM-----YI-----LRTVPERAIDTDSNLTWEG 909
Db 549 NNNKKQBERMLKQIENLOETETQLRNELEYVREELKQKRDVCKLDKSEENCNNLRKQV 608
QY 910 MIQMAFHIALGLLEKQOQKAPBEEVTFDYHKASRLGSSAMNT-QMLLEKLGIPOL 968
Db 609 ENKNKY-----IEELOQENKALKKKGTAE-----SKQLNVEIKVANKLE--LEL 650
QY 969 EQOKDMITWLOMFDVTKRLREKSCSLIVATTSSESINKNDEITHDKEAERKKAAEARL 1028
Db 651 ESAKOKFGEITTYQ--KEIEDK-----KISEENLEEVEKA-KVIADBAVL 695
QY 1029 HR-----QKIMASALQKNTIETHKMYDNTSEMPGKEDST--MEEESTPAVSDYSRI 1080
Db 696 QKEIDKRCQHKTAEMVAL-----MEKHQYDKIIERDSELGLYKSKEQESSLASLEI 751
QY 1081 ALGPRGRPSVTEKEVLTCILOEQEVEVKIENNAWLSACVQKSTALTQHRGKPIELSGEA 1140
Db 752 ELSNLKAELLSVKKQLE-----IEREKEKLKREAK-----ENTATLKEKKKTKQTF--- 799
QY 1141 LDPFLWDPDLATYGTGSGCHVHAWCVQKYEAV--QLSSQORIHVDL-FDLESEGYLC 1197
Db 800 ---LLETPE-----IYWKLDKAVPSQTVSRNFTSVDHGISKOKRDLWM 840
QY 1198 PLCKSLCNTVIP-----IIPLOPKINSEN 1222
Db 841 TSAKNTLSTPLPKAYTKVTPKPKLOOREN 870

RESULT 10
US-08-508-836A-8
; Sequence 8, Application US/08508836A
; Patent No. 5777093
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-Telangiectasia Gene
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/508,836A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-313 (TAU)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-508-836A-8

Query Match 1.4%; Score 133.5; DB 1; Length 3056;
Best Local Similarity 17.9%; Pred. No. 0.061;
Matches 260; Conservative 196; Mismatches 470; Indels 529; Gaps 68;
QY 319 SWNKMINSYSDPQIFCQACLR-----EPDSENP-----CL---ISRLMLWD 359
Db 1857 SWRNLLSTHV---OGFFTCLRHFSQTSRSTTPANLDSSEHFFRCCLDKKSQRTMLAV 1912
QY 360 AKLYKGARKILHELIFS-SFEMEMEKYL-----FAMERVKYKQIQKEYISDD 407
Db 1913 VDMRRQKRSPSSGTFPDAFWLDLNLLEVAKVQSCAAHFTALLYAEIY--ADKSMDDQ 1970
QY 408 HDRSISITALSVMQFTVPTLARHLEBQNVTSVITETLLEVLPEYLDNRNKNFQGYSD 467
Db 1971 EKRSAPAEES-QSTTSSISEKSKEETGI--SLQDLLEIYRSIGEPDSLYGCGG--- 2023
QY 468 KLGRVYAVICDLKYLISKPTIWERLRMQPLEGFRSLFKILTCMQGHEETRRVQCO-HI 526
Db 2024 --GKMLQPIITFLR--TYEHEAMWGKALVTYDLE--TAIPSTRQAGIIQALQNLGLCHI 2076
QY 527 EVDPDWEAATAIOMQLKNILLMFOEWACADELLIVA-----YKECHKAVMRC--STSFI 579
Db 2077 -----LSVYLAGLDYENKDWCPLEELHYQAARNMOWDHCTSVSKEVEGTSYH 2125
QY 580 SSKTVVQSCG-----HSLETQSYRVSDLVSIHPLSRTLAGLHVRLSRGAVSRL 631
Db 2126 ESNALQSLDRDREFSTFYESLKYARVKEVEMCKRSL---ESVSYLPTLSRLQAIGEL 2182
QY 532 HEFVSFEDFQV-----EVLVEYPLRCLVLVAOVVAEMRRNGLSLSQVYVQDV--- 681
Db 2183 ESTGELFSRVTHRQLSEVYIK-----WOKHSQLLKDSDFSQEPIMAL 2226
QY 682 -----KCREMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAEAFNK 723
Db 2227 RTVILEILMEKENDNSORECIKILTKHLVELSI-----LARTFN 2267
QY 724 T-ISTKDQDLIKOYNTL-----IEEMQLVLIYVGRYVPGVGNVTKKEVTMRIIH 774
Db 2268 TQLPERAIFOIKQYNSVSCGVSEMQLEE-AQVF-----WAKREOSLALSIL 2312
QY 775 LLCIEPMPHSAIAKN-----LPENNETGLNVINKVATFKPGVSGHGYV- 821
Db 2313 KOMIKKLDASCAANNPSLKLTYTECLRVCGNWLAECLTENPAVIMQTYLEKAVEVAGNYD 2372
QY 822 -----ELKDESLEKDFNMFYHYHSKTOHSAEHMOK-----KRRKQJ- 857
Db 2373 GESSDELNRNGMKKAF--LSLARFSDTYQRIENYMKSEFENKQALLKRAKEEVLRLRHK 2431
QY 858 -----NKDE-----ALPPPPPEFCPAFSKVINLLNCDIM-----MYILRTV 894
Db 2432 IQTNRYTVKQVRELELDELALRALKEDRKRFELCKA---VENYINCLLSGEEHDMVFL- 2487
QY 895 FERAIDTDSNLWTEGMLQMAFHIALGLLEKQOQKAPBEEVTFDF-----YHKASRLGS 950

Db 2488 -----CSWLLEN-----GVSEVNGMMKRGKMPITYKFLPLMYQLAARMGT 2529
Qy 951 SAMTIOMLEKLGIPLEGOKMITWILQMFDTVKRLREKSLIVATPSSSESIKNDI 1010
Db 2530 KMG-----GLGFHEVLNLLISRI-----SMDHPHHTLFIILANAN-----RDEF 2571
Qy 1011 THDKEKAERKR-----KAEAA-----RLHRQKIMQMSALQKNFI--- 1045
Db 2572 LTKPEVARRSRITKNVPKQSSOLDREDTEAANRIICTIRSRPQMVRSVEALCDAIILA 2631
Qy 1046 -----ETHKLMYDNTSEMPKEDSIMEESTPAV-----SDYS-----R 1079
Db 2632 NLDATQMTORGINIPADQITLKNLEDDVVVPTMEIKVDHTGEYGNLVITQSFKAERF 2691
Qy 1080 IALGPKRGPSVTEKVLTCI-----LCQEEQEVKTEN 1111
Db 2692 LAGG-----VNLPIIDCVSGDKERRQLVKGRDDLQDAVMQOVQMCNTLILQRTET 2745
Qy 1112 NAMVLSACVOKSTALTOHRG-----KPIELSGEALDPLFMDPLAYGTYTSCGHVM 1163
Db 2746 RKRKLITICTYKVPPLSQRSGLVWCTGTVP-----GEF-----LVNNEDEGAHKRYRPNDFSFA 2799
Qy 1164 HAVCWQKYFEAVQLSSOORIHVDLFDESGLYLCPLCKSLCNTVPIPILOPQKINSENA 1223
Db 2800 Q--COKMMEVQKKSFEEKVEVFM-----DVCNQFPVF-----RY 2833
Qy 1224 DALAQLTLARWITVLA-----RISGYNIRHAKGENPIPIFFNOGMDSTLEFHSI 1275
Db 2834 FCMKEFLDPAWFKRLAYRSVATSSIVGIL-----GLGDRHVQ----- 2874
Qy 1276 LSGVSESIKYSIKEMVILFATYIRIGLVKPPDERDP-----RVPMILT 1321
Db 2875 -----NILNEQSAELVHIDLGVAFEQKILPTPETVPRLTRDIDVGNGITGVGVF 2927
Qy 1322 WSTCAFTIOAIEN-----LLGDEGKPLFGALQNRHNGLKALMFAVQRTCP 1370
Db 2928 RRCEKTEMEVRNSQETLLATIVELLYD---PLF---DWTMNPALKALY---LQOR---P 2974
Qy 1371 QVLIOKHLVRLSVVPLNPKSEDPCLLSIDLFLVLCVAVLAPFLSLYWDVDPVLOPSSVS 1430
Db 2975 EDETELH-----PTLNADOEQCRNL-----SDID 2999
Qy 1431 SSYNHLYFLHITMAHMLQILLTVDTGPLA---QVQ---EDSEAHSSASSFFAEISQYT 1484
Db 3000 QSPDKVAERVLMLRLOELK---GVEEGTVLSVGQVNLIIQOAIIDPNKLSLRF----- 3049
Qy 1485 SGSGICDIPGWYLV 1499
Db 3050 -----PGWKAWV 3056

RESULT 11

US-08-629-001A-3
; Sequence 3, Application US/08629001A
; Patent No. 5858661
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5858661thwestern Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/629,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2290,00032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-629-001A-3

Query Match 1.4%; Score 133.5; DB 2; Length 3056;
Best Local Similarity 17.9%; Pred. No. 0.061;
Matches 260; Conservative 196; Mismatches 470; Indels 529; Gaps 68;

Qy 319 SWNMKMSYSSDFRQIFCOACILRE-----EPDSENP-----CL--ISRMLWD 359
Db 1857 SWRNLLSTHV---QGFTTCLRHFSQTSRSTTPANLDSEHFFRCCLDKKSORTMLAV 1912
Qy 360 AKLYGARGKILHELIFS-SFFMEMEYKKL-----FAMEFVKYKQLOKEYSDD 407
Db 1913 VDMYMRQRPSSTGFENDAFWLDLNVLEKVAQSCAAHTALLYAEIY--ADKKSDDQ 1970
Qy 408 HDRSITISALSVQMTPTVPTLARHLIEBQNVISVITLLEVLPEYLDNRNKNFQGYSD 467
Db 1971 EKSLAFEGS-QSTTSSLSSEKSEETGI--SLQDILLLEIYRSIGEPDLSYCGG--- 2023
Qy 468 KLGRVYAVICDLKYLISKPTTWTERLRQWLEGFSFLKILTCMQGMEIRQVQ--HI 526
Db 2024 --GKMLQPIITRL--TYHEAMWGKALVTYDLE--TAIPSSSTRQAGITQALQNLGCHI 2076
Qy 527 EYDVPDWEAAIAIQOMLNILLMFOEWCACDEELLVA-----YKECHKAVMRC-STFSI 579
Db 2077 -----LSVLKGLDYENKDWCPLEELHYQAANRNMOWDHCTSVSKEVEGTSYH 2125
Qy 580 SSKTWTWQSCG-----HSLETKSYRVSDELVSIHPLSRTLGLHLVRLSRLGAVSRL 631
Db 2126 ESLYNALQSLRDREFSTFYESLKYARVKEVEEMCKRSL---ESVSYLPTLSRLQAI 2182
Qy 632 HEFVSFEFQV-----EVLVEYPLCLVLVAQVAVMRRNGLSLISQVFFYQDV---- 681
Db 2183 ESIGELFSRSVTHROLSEVIK-----WQKHSOLLKSDSFSQEPIMAL 2226
Qy 682 -----KCREEMYDKDITIMIQIGASLMDPNKFLLLVLQRYELAEAFN 723
Db 2227 RTVLEILMEKEMDMSORECIKDILTKHLVLSI-----LARTFKN 2267
Qy 724 T-ISTKDQDLIKOYNTL-----TEEMLQVLIYIVGERYPGVGNVTKKEEYTMREI 774
Db 2268 TQLPERAIFQIKOYNSVSGVSEWQLEE-AQVF-----WAKKEQSALSL 2312
Qy 775 LCIEPMPHSATAKN-----LPENENNETGLENVINKVATFKKPGVSGHGVY- 821
Db 2313 KOMIKKLDASCAANNPSLKITVTECLRVCGNWLAECTLENPAVIMQTYLEKAVEAGNYD 2372
Qy 822 -----ELKDESLKDFNMYYHYHYSKTOHSAEHMQ-----KRRKQE----- 857
Db 2373 GESSDELNGMKAF--LSLARFSDTOYQRIENYMKSEFENKQALLKRAKEEYGLREHK 2431
Qy 858 -----NKDE-----ALPPPPPEFCFAPAFSKVINLLNCIDIM-----MYILRTV 894
Db 2432 IQTRNYTVKQVQRELELDELALRALKEDKRFLCKA---VENYINCLLSGHEHDMWVRL- 2487

QY	895	FERAIDTSDNLWTEGMLQWAPHILALGILLEEKQLOKAPEEVTFDF----	YHKASRLGS	950
Db	2488	-----CSLWLENS-----GYSEVNGMMKRGDMKIPYKFLPLMYQLAARMGT	2529	
QY	951	SAMNIQMLLEKLGJPOLEGOKDMITWILQFDMTVKRLREKSCULIVATTSGSESINDEI	1010	
Db	2530	KMMG-----GLGFHEVLNLLISRI-----SMDPHHHTFTILALANAN-----RDEF	2571	
QY	1011	THDKAEAKRR-----KAEAA-----RLHROKIMAQMSALQKNFI----	1045	
Db	2572	LYKPEVARSRITKNVPKQSSOLDEDRTEAANRLIICIRSRPQWRSVEALCDAYILLA	2631	
QY	1046	-----ETHKLMYDNTSEMPCKEDSIMEESTPAV-----SDYS-----R	1079	
Db	2632	NLDATQWKTKRGINIPADQPIITKLKNLEDVVVPTMEIKVDHTGEYGNLVTIQSKRAEFR	2691	
QY	1080	IALGPKRGSPVTEKEVLTCI-----LCOEQEVEKVIEN	1111	
Db	2692	LAGG-----VNLPKIITDCVSDGSKERRQLVKGRDLRQDAVMQOVQMCNTLLQRNTET	2745	
QY	1112	NAMVLISACVOKSTALTOHRG-----KPIELSGEALDPLFMDPDPLAYGTGYTSCGHVM	1163	
Db	2746	RKRKLITICTYKVPVLSQRSGVLEWCTGVPPI---GEF---LVNNEDGAHKRYRPNDFSAF	2799	
QY	1164	HAVCMQKYPEAVOLSSQORIHVDLFLESGEVLCPLCKSLCNTVPIPIIPLQPKINSENA	1223	
Db	2800	Q---CQKKMEVOKKSFEEKYEVFM-----DVCQNFQPVF-----RY	2833	
QY	1224	DALAQLLTLARMIQTVLA-----RISGYNIRHAKGENPIPIFFNQMGMDSTLBFHSI	1375	
Db	2834	FCMEKFLDPAIWFEKRLATRSVATSSIVGIL-----GLGDRHQVQ---2874		
QY	1276	LSFGVESTKYSNSTKEWVILFATTYIRIGLUKVPDPERDP-----RVPMLT	1321	
Db	2875	-----NILINEQSAELVHIDLGVAFBOGKILPTPTVPRLTRDIVDGMGITGVEGVF	2927	
QY	1322	WSTCAFTIOAIEN-----LLGDGKPLFGALQNRQHNGKLLKALMQFAVAQRITCP	1370	
Db	2928	RRCCKETMEVMRNSOETLLTIVIVLLYD---PLF---DWTMNPALKALY---LQOR---P	2974	
QY	1371	QVLIQKHLVRLLSVVLPNPKISEDPTCLLSIDLFLHVLCVAVLAPSLXWDDPVDLQPSVVS	1430	
Db	2975	EDETELH-----PILNADDOECKRNL-----SDID	2999	
QY	1431	SSYNHLYLFHLITMAHMLQIILLTVDITGLPLA---QVQ---EDSEEAHSASSFFAEISQYT	1484	
Db	3000	QSFDKVAERVLMLRLOEKLK---GYEEGTVLSGVGOVNLIIQQAIDPKNLSRLF-----	3049	
QY	1485	SGSIGCDIPGWYLW	1499	
Db	3050	-----PGWKAWV	3056	

RESULT 12

```

RESOLUTION 1-2
US-08-874-266-2
; Sequence 2, Application US/08874266
; Patent No. 5955279
; GENERAL INFORMATION:
; APPLICANT: Gatti, Richard A.
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA: MUTATIONS IN THE ATM GENE
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```


Db 2373 GESSDELNGCKMAF--LSLARFSDTQYORIENYMKSEFFENKQALLKRAKEEVGLLREHK 2431
Qy 858 -----NKDE-----ALPPPPPPFCFAFSKVINLLNCDIM-----MYILRTV 894
Db 2432 IOTNRYTVKVQRELEDELALRALKEDKRFLCKA---VENYINCLLSGEEHDMWVFL- 2487
Qy 895 FERAIDTDSNLWTEGMLQMAFHIALGLLEKQOLQKAPEEVTFDF-----YHKASRLGS 950
Db 2488 -----CSLWLENS-----GVSEVNGMMKRDGMKIPTYKFLPLMYQLAARMGT 2529
Qy 951 SAMNTQMLEKLGIPQLEGOKMITWILQMFDTVRLREKSLIVATTSSSESIKNDIEI 1010
Db 2530 KMMG-----GLGFHEVLNLLSRI-----SMDPHHTLFIILALANAN-----RDEF 2571
Qy 1011 THDKEAERKR-----KAEAA-----RLHROKIMQMSALOKNFT--- 1045
Db 2572 LTKPEVARRSRITKNVPKQSSQLEDDETEAANRIICTIRSRPOMVRSVEALCDAYIILA 2631
Qy 1046 -----ETHKLMYDNTSEMPGKEDSIMEESTPAV-----SDYS-----R 1079
Db 2632 NLDATQWKTKRGKINIPADQPIITLKNLEDDVVVPTMEIKVDHTGEYGNLVTIOSFKAEFR 2691
Qy 1080 IALGPKRGPSVTEKVLTCI-----LQOEBOEVKIEIN 1111
Db 2692 LAGG-----VNLPKIITDCVSDGKERRQLVKGRRDLRQDAVMQOVQCMNTLLQRNTET 2745
Qy 1112 NAMVLSACVKSTALTOHRG-----KPIELSGEALDPLFMDPLAYGTYTSCGHVM 1163
Db 2746 RKRKLITCTYKVPLSORSQVLEWCTCTVPI---GEF---LVNNEGAKHRYRPNDFSAP 2799
Qy 1164 HAVCWOKYFEAVOLSSOORIHVDLFDESGEYLCPLCKSLCNTVPIPIPLQPKQKINSENA 1223
Db 2800 Q--CQKMMVEQKKSPEEKEVEN-----DVCONQPVF-----RY 2833
Qy 1224 DALAQTLTLARNIQTULA-----RISGYNIRHAKGENPIPIPFNOGMDSTPLEFHSI 1275
Db 2834 FCMKEFLDPAIWFKEKRLAYTRSVATSSIVGYL-----GLGDRHVQ--- 2874
Qy 1276 LSFGEVSSIKYSIKEMVILFATTIYRIGLKVPPDERP-----RVPMLT 1321
Db 2875 -----NILINQSAELVHIDLGVAFEQGLKILPTPTVFRRLTRDIDVGNGITGVSGVF 2927
Qy 1322 WSTCAFTIOAIEN-----LLGDEGKPLFGALQNRQHNGLKALMOFAVAORITCP 1370
Db 2928 RRCCEKTEVWRNSQETLLTIVEVLLVD---PLF---DWTMNPALKY---LQOR---P 2974
Qy 1371 QVLIQKHLVRLSVLVPNIKSEDTPCLLSIDLHFVLVGAVLAPPSLYWDVDPVQLQPSVS 1430
Db 2975 EDETELH-----PTLNADDOECKRNL-----SDID 2999
Qy 1431 SSYNHLYFLHITMAHMLQILLTVDTCPLA---QVQ---EDSEAHSSAFFAEISQYT 1484
Db 3000 QSFQKVAERVMRLQEKLK---GVEEGTIVSGVGQVNLIIQQAIDPKNLSRLF----- 3049
Qy 1485 SGSGICDIPGWYLMW 1499
Db 3050 -----PCWKAWV 3056

RESULT 13

US-08-642-274D-3
; Sequence 3, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
; TITLE OF INVENTION: SCREEN FOR A PARTIAL A-T PHENOTYPE
; FILE REFERENCE: 22900033
; CURRENT APPLICATION NUMBER: US/08/642,274D
; CURRENT FILING DATE: 1996-05-03
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3

; LENGTH: 3056

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-642-274D-3

Query Match

Best Local Similarity 1.4%; Score 133.5; DB 4; Length 3056;

Matches 260; Conservative 196; Mismatches 470; Indels 529; Gaps 68;

Qy 319 SWMKIMSYSSDFRIFCOACURE-----EPDSENP-----CL---ISRLMLWD 359

Db 1857 SWRNLLSTHV---QGFETCLRHFSQTSRSTTPANLSDSEHFFRCRCDKKSQRTMLAV 1912

Qy 360 AKLYKGARKILHELIFS-SFFMEMEYKLI-----FAMEFVKYKQLOKEVISDD 407

Db 1913 VDYMRQRQRPSSGTFNDAFWLDLNLVLEKVAQSCAAHFTALLYAEIY--ADKKSMDDQ 1970

Qy 408 HDRSITALSVMQFTVPTLARHLIEEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQ 467

Db 1971 EKRSIAFEES-QSITISLSEKSEETCI--SLQDLLLEIYRISIGEPSDLYCGG--- 2033

Qy 468 KLGRYVAVICLKYILISKPTIWERLMOFLEGFRSFLKILTCMOGMEBEIRQVGO-HI 526

Db 2024 --GKMLQPIITLR--TYHEAMWGKALVTYDLE---TAIPSSTRQAGIIQALQNLGLCHI 2076

Qy 527 EVDPPWEAAIAIOMOLKNILLMFQEWACDDEELLVA-----YKECHKAVMRC-STSF 579

Db 2077 -----LSVYLKGLDYENKDMCPLEELHYQAAWRNMOWMDHCTSVSKEVEGTSYH 2125

Qy 580 SSSKTVVQSCG-----HSLETKSRYVSEDLVSTHPLSRTLGLHLVRLSRLGAVSRL 631

Db 2126 ESNALQSLDRDREFSTFYESLKYARKEVEEMCKRSL---ESVSYLPTLSRLQALIGEL 2182

Qy 632 HEFVSFDFQV-----EVLVEYPLRCLVLVAQVVAEMMRNGLSLISQVIFYQDV--- 681

Db 2183 ESIGELFSRVTHRQLSEVYIK-----WOKHSQLLKDSDFSQEPIAL 2226

Qy 682 -----KCREMYDKDIIMLQIGASLMDPNKFPILLVLQRYELAEAFNK 723

Db 2227 RTVILEILMEKEMDSQRECIKDILTTLHLVELSI-----LARTFKN 2267

Qy 724 T-ISTKDODLKOYNTL-----IEEMLOVLIYIVGERVPGVGNVTKEEYTMREIHH 774

Db 2268 TQLPRAIFQIKQYNSVSCGVSEWOLEE-AQVF-----WAKKEQSLSLSIL 2312

Qy 775 LLCIEPMPHSAIAKN-----LPENENNETGLENVINKVATFKKPGVSGHGVY- 821

Db 2313 KOMIKKLDASCAANPSLKLTVTECLRVCGNMLAETCLENPAVIMQTYLEKAVEVAGNYD 2372

Qy 822 -----ELKDESLLKDFNMVYFYHYSKTOHSAEHMQK-----KRRKQE----- 857

Db 2373 GESSDELNGCKMAF--LSLARFSDTQYORIENYMKSEFFENKQALLKRAKEEVGLLREHK 2431

Qy 858 -----NKDE-----ALPPPPPPFCFAFSKVINLLNCDIM-----MYILRTV 894

Db 2432 IOTNRYTVKVQRELEDELALRALKEDKRFLCKA---VENYINCLLSGEEHDMWVFL- 2487

Qy 895 FERAIDTDSNLWTEGMLQMAFHIALGLLEKQOLQKAPEEVTFDF-----YHKASRLGS 950

Db 2488 -----CSLWLENS-----GVSEVNGMMKRDGMKIPTYKFLPLMYQLAARMGT 2529

Qy 951 SAMNTQMLEKLGIPQLEGOKMITWILQMFDTVRLREKSLIVATTSSSESIKNDIEI 1010

Db 2530 KMMG-----GLGFHEVLNLLSRI-----SMDPHHTLFIILALANAN-----RDEF 2571

Qy 1011 THDKEAERKR-----KAEAA-----RLHROKIMQMSALOKNFT--- 1045

Db 2572 LTKPEVARRSRITKNVPKQSSQLEDDETEAANRIICTIRSRPOMVRSVEALCDAYIILA 2631

Qy 1046 -----ETHKLMYDNTSEMPGKEDSIMEESTPAV-----SDYS-----R 1079

Db 2632 NLDATQWKTKRGKINIPADQPIITLKNLEDDVVVPTMEIKVDHTGEYGNLVTIOSFKAEFR 2691

```
QY 1080 TALGPKRPSVTEKVLTCI-----LCQEEQEVKIN 1111
:| | | | | | | | | | | | | | | | | | | | |
Db 2692 LAGG-----VNLKRIIDCVSGDKERRQLVKGRDDLRODAVMQVQFQMCNTLLQRNTET 2745
:| | | | | | | | | | | | | | | | | | | | |
QY 1112 NAMVLSACVOKSTALTOHRG-----KPIELSGEALDPLEMDPLAVGTGTGSCGHVM 1163
:| | | | | | | | | | | | | | | | | | | | |
Db 2746 RKRKLITICTYKVVPSQSRGVLEWCTGTVP-----GEF-----LVNEDGAHKRYRPNDFSAP 2799
:| | | | | | | | | | | | | | | | | | | | |
QY 1164 HAVCMQKYFEAVQLSSQORIHVDLFDESGEYLCPLCKSLCNTVPIPIPLQPKINSENA 1223
:| | | | | | | | | | | | | | | | | | | | |
Db 2800 Q--CQKMEVQKKSFEKYEYFM-----DVCONFQPVF-----RY 2833
:| | | | | | | | | | | | | | | | | | | | |
QY 1224 DALAQLLTLARWITVLA-----RISGYNIRHAKGNPIPIFFNOGMDSTLEFHSI 1275
:| | | | | | | | | | | | | | | | | | | | |
Db 2834 FCMERFLDPAIWFKRRLAYTRSVATSSIVGYIL-----GLGDRHVO----- 2874
:| | | | | | | | | | | | | | | | | | | | |
QY 1276 LSFVGESSIKYSNKEMVILFATYIRIGLKVPPDERP-----RVPMLT 1321
:| | | | | | | | | | | | | | | | | | | | |
Db 2875 -----NILINEQSAELHIDLVAFEOGKILPTPETVPRFLTRDIVDGMGITGVEGVF 2927
:| | | | | | | | | | | | | | | | | | | | |
QY 1322 WSTCAFTQAIEN-----LLGDEGKPLFGALQNRHNGKALMOFAVAQRITCP 1370
:| | | | | | | | | | | | | | | | | | | | |
Db 2928 RRCCEKTEVMRNSQETLLTIVEVLLYD---PLF-----DWTNPLKALY---LQOR---P 2974
:| | | | | | | | | | | | | | | | | | | | |
QY 1371 QVLIQKHLVRLLSVVLNPKTSEDTPCLLSIDLFHVLVGAVLAPPSLYWDDVDLPSSVS 1430
:| | | | | | | | | | | | | | | | | | | | |
Db 2975 EDETELH-----PTLNADDQECKRNL-----SDID 2999
:| | | | | | | | | | | | | | | | | | | | |
QY 1431 SSYNNHLYFLHITMAHMLIOLLTVDTLGLPLA---QVO---EDSEEAHSASSFAEISQVT 1484
:| | | | | | | | | | | | | | | | | | | | |
Db 3000 OSFDKVAERVLMLQEKKLK---GVEEGTVLSGGVGNLLIQOAIIDPKNLSRLF----- 3049
:| | | | | | | | | | | | | | | | | | | | |
QY 1485 SGSIGCDIPGWYLVW 1499
:| | | | | | | | | | | | | | | | | | | | |
Db 3050 -----PGWKAVW 3056
:| | | | | | | | | | | | | | | | | | | | |
```

RESULT 14

```
US-08-952-127-3
; Sequence 3, Application US/08952127
; Patent No. 6211336
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: ATAXIA-TELANGIECTASTA GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 6211336thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,127
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 2290.00029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-539-5050
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-952-127-3
```

Query Match 1.4%; Score 133.5; DB 4; Length 3056;

Best Local Similarity 17.9%; Pred. No. 0.061; Matches 260; Conservative 196; Mismatches 470; Indels 529; Gaps 68;

```
QY 319 SWNMKIMSYSDFRQFCQACLR-----EPDSENP-----CL---ISRLMLWD 359
:| | | | | | | | | | | | | | | | | | | | |
Db 1857 SWRNLLSTHV---QGFFTSCLRHSQTSRSTTPANLDSESEHFRCCLDKKSQRTMLAV 1912
:| | | | | | | | | | | | | | | | | | | | |
QY 360 AKLYKGARKILHELIFS-SFFMEMEYKKL-----FAMEFVKYKQLOKYEYISDD 407
:| | | | | | | | | | | | | | | | | | | | |
Db 1913 VDYMRQRKRPSSGTIFNDAFWLDLNYLEVAQVSCAAHFTALLYAEIY---ADKKSMDQ 1970
:| | | | | | | | | | | | | | | | | | | | |
QY 408 HDRSISITALSVMFTVPTLARHLIEQNVISVITETLLEVLPEVLDNRNKNFQGYSD 467
:| | | | | | | | | | | | | | | | | | | | |
Db 1971 EKSLAPEES-QSTTSSISEKSEETGI---SLQDLLLEYIRSIGEPSLYGCGG----- 2023
:| | | | | | | | | | | | | | | | | | | | |
QY 468 KLGKVAVICDLKYLISKPTIWERMQFLEGFRSLKILTCMOGMEERQVQ-HI 526
:| | | | | | | | | | | | | | | | | | | | |
Db 2024 --GKMLQPIPLR--TYEHEAMGKALVTYDLE---TAIPSTRQAGIIQALNGLCHI 2076
:| | | | | | | | | | | | | | | | | | | | |
QY 527 EVDPDWEAATAIOMKLNILLFQEWACADELLVA-----YKECHKAVMRC-STSEI 579
:| | | | | | | | | | | | | | | | | | | | |
Db 2077 -----LSVYLGKLDYENKDWCPLELHYQAARNQMDHCTSVSKEVEGTSYH 2125
:| | | | | | | | | | | | | | | | | | | | |
QY 580 SSSKTVVQSG-----HSLETKSVRSVSDLVSIHLPLSRTLGLHVLRLSRLGAVSRL 631
:| | | | | | | | | | | | | | | | | | | | |
Db 2126 ESYNALQSLRDREFFTFYESLKYARKEVEEMCKRSL---ESVSYLPTLSRQAIGEL 2182
:| | | | | | | | | | | | | | | | | | | | |
QY 632 HEFVSFEDFQV-----EVLVEYPLRCLVLVAQVVAEMRRNGLSLISQVYVQDV----- 681
:| | | | | | | | | | | | | | | | | | | | |
Db 2183 ESIGELFSRSVTHRLSEVIK-----WKHSQLLKDSDFQFPIHAL 2226
:| | | | | | | | | | | | | | | | | | | | |
QY 682 -----KCREMYDKDIIMLQIGASLMDPNKFLLLVLQRYEALAEANK 723
:| | | | | | | | | | | | | | | | | | | | |
Db 2227 RTVILEITMEKEMDNSQRECIKDLTKHLVELSI-----LARTEN 2267
:| | | | | | | | | | | | | | | | | | | | |
QY 724 T-ISTRDODLKOYNTL-----IEMLQVLIYIGERVVPGVGNVTKKEVTMRLEIH 774
:| | | | | | | | | | | | | | | | | | | | |
Db 2268 TQLPERAIFQIKQYNSVSCGVSEWQLEE-AQVF-----WAKKEQSLSIL 2312
:| | | | | | | | | | | | | | | | | | | | |
QY 775 LUCIEPMPHSAIAKN-----LPENENNTEGLENVINKVATFKKPGVSGHYV- 821
:| | | | | | | | | | | | | | | | | | | | |
Db 2313 KOMIKKLADSCAANNPSLKLTYTECLRVCGNWLAECLENPAVIMQTYLEKAVEAGNYD 2372
:| | | | | | | | | | | | | | | | | | | | |
QY 822 -----ELKDESCLKDFNMFFYHYSKTHSKAEHMOK-----KRRKOE----- 857
:| | | | | | | | | | | | | | | | | | | | |
Db 2373 GESSDELNRGKMKAF-LSLARFSDTOYRTENYMKSEFENKQALLKRAKEEVLGREHK 2431
:| | | | | | | | | | | | | | | | | | | | |
QY 858 -----NKDE-----ALPPPPPPFCFAPSKVINLLNCDIM-----MYILTV 894
:| | | | | | | | | | | | | | | | | | | | |
Db 2432 IQTNRYTVKQRELELDELALRALKEDRKRFCKA---VENYINCLISGEEDHMMVFL- 2487
:| | | | | | | | | | | | | | | | | | | | |
QY 895 FERAIDTDSNLWTEGMLQMAFHILALGLLEKQOLOKAPEEVTFDF-----YHKASRLGS 950
:| | | | | | | | | | | | | | | | | | | | |
Db 2488 -----CSLWLENS-----GVSEVNGMKRDKGMKIPTYKFLPLMYQLAARMGT 2529
:| | | | | | | | | | | | | | | | | | | | |
QY 951 SAMNIOMLLEKLGIPQLEGQKDMITWILQMFDTVKRLREKSKCLIVATTSSESTKNDI 1010
:| | | | | | | | | | | | | | | | | | | | |
Db 2530 KMMG-----GLGFHEVLNLLISRI-----SMDPHPHHTLFLILALANAN----RDEF 2571
:| | | | | | | | | | | | | | | | | | | | |
QY 1011 THDKKAERKR-----KAEAA-----RLHRQKIMAOMSALOKNFI--- 1045
:| | | | | | | | | | | | | | | | | | | | |
Db 2572 LTKPEVARRSRITKNVPKQSSQLDEDRTEAANRLICTRSRRPQWRSVEALCDAYILA 2631
:| | | | | | | | | | | | | | | | | | | | |
```

```
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-952-014C-3
```

	Query Match Best local similarity Matches	1.4%; 17.9%; 260;	Score 133.5; Pred. No. 0.061; Conservative 196;	DB 4; Mismatches 470;	Length 3056; Indels 529;	Gaps
Qy	319 SWMNKIMSYSDPFIQCOACURE-----EPDSENP-----CL---ISRILMLWD	359	: :: - - - - - QGFTTSCLRHFHSQTSSRTTPANLDSSEHFFPCCLDDKKKSORTMLAV	1912	: : :	
Db	1857 SWRNLLSTHV-----QGFTTSCLRHFHSQTSSRTTPANLDSSEHFFPCCLDDKKKSORTMLAV	1912	: : :			
Qy	360 AKLYKGARKHLHELIFS-SFFMEMEYKKL-----FMFEVVKYIKQLQEYISDD	407	: : : : ::: : : : : : : : : :			
Db	1913 VDMRMRQRPPSGGTIFNDAFWLDNLNVLEVAKVAAQSAAHTALLYAEIV--ADBKSMDDQ	1970				
Qy	408 HDRSISIATLSVMQVTPTFLARHLEEQNVI SVITETILLEVLPEYLDRNNKNFNFGYSQD	467	: : : : : : : : : : : :			
Db	1971 EKRSIAFEEGS-QSTTSISLKSKKEETGI--SLQDLLEIYRSIGEPSDSLVCGG---	2023	: : : : : : : : : : : :			
Qy	468 KLGRVVAVICDLKYILISKPTTWTERLRMQELGEPRFSFKILTMOGMEEIRROVGQ-HI	526	::: : : : : : : : : : :			
Db	2024 --GKMLOPIITUR--TYEHFAWMGRALVTYDLE---TAIPSTRQAAGIIQALONGLCHI	2076	: : : : : : : : : : : :			
Qy	527 EVDPDWEAATAIQMKLNILLMFQQWCACDEELLIVA-----YKECHKAVMRC-STSF I	579	:			
Db	2077 -----LSVYLKGLDYENKWCPLEELHHYAARNMWODHCTSVSVEVEGTSYH	2125	:			
Qy	580 SSSKTWVSOCG-----HSLETKSYRVSEDLVSI THPLSR TL AGLHVR LSR LGAVSR L	631	:			
Db	2126 ESLYNALOSLRDPREFSTFYESLUK Y AR VK EEMCR KSL ---ESVSYLP T TS LR LOA IG EL	2182	:			

[illegible]

```
QY 1011 THDKKAEKRK-----KAEAA-----RLHQKIMAOMSALQKNFI--- 1045
Db 2572 LTKPEVARRSRTIKNVKQSSOLDDETRTEAANRIICTIRSRPQMVRSVEALCDAYIILA 2631
QY 1046 -----ETHKLMVDNTSEMPCKEDSIMESTPAV-----SDYS-----R 1079
Db 2632 NLDAQWKTORKGINIPADQPTIKLNLEDVVVPTMEIKVDHTGEYGNLVTIQSFKAEFR 2691
QY 1080 IALGPKRGPSTEKEVLFCI-----LCQEEQEVKNIEN 1111
Db 2692 LAGG-----VNLKIIDCVSGDKERRQLVKGRDDLQDAVMQOVQFQMCNTLLQRNTET 2745
QY 1112 NAMVLSACVQKSTALTQHRG-----KPIELSGEALDPLFMDPDLAYGTYTSCGHVM 1163
Db 2746 RKRKLITICTYKVVPLUSQSGVLEWCTGTVP-----GEF-----LVNNEGAKHRYRPNDFSAF 2799
QY 1164 HAVCQKYFEAVOLSSQORIHVDLDESGEVLCPLCKSLCNTVPIIIPLOPKINSENA 1223
Db 2800 Q--COKMMEVQKKSFEKYEVM-----DVCFQFQV-----RY 2833
QY 1224 DALAQLLTARWIQVLA-----RISGYNIRHAKGENPIPIFNQGMGDSSTLEFHSI 1275
Db 2834 FCMKFLDPAIWFKEKRLAYTRSVATSSIVGYL-----GLGDRHVQ----- 2874
QY 1276 LSFGEVSSIKYSNIKEMVILFATTIYRIGLKVPDERDP-----RVPMLT 1321
Db 2875 -----NILNEQSAELVHIDLGVAFEGQKILPTETVPFRLTRDIVDGMGITGVEGVF 2927
QY 1322 WSTCAFTIOAIEN-----LLGDEGKPLFCALONROHNGKALMQFAVAORITCP 1370
Db 2928 RRCCKTWEVRNRSOETLLTIVEVLLYD---PLF---DWTNPLKALY---LQOR---P 2974
QY 1371 QVLIQKHLVRLLSVVLPNIKSEDTPCLLSIDLHVLVGAVLAFPSLYWDDPDVDLPSSVS 1430
Db 2975 EDETELH-----PTLNADDQCKRNL-----SDID 2999
QY 1431 SSYNHLYLFLHITMAHMLQILLTVDTLPLA---QVQ---EDSEAHSSASSFFAEISQYT 1484
Db 3000 QSFDKVAERVLMLRQEKLK---GVEEGTVLSVGGQVNLIIQQAIDPKNLSRLF----- 3049
QY 1485 SGSIGCDIPGWYLVW 1499
Db 3050 -----PGWKAWV 3056
```

Search completed: February 28, 2002, 10:02:00

Job time: 297 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 09:56:13 ; Search time 35.89 Seconds
(without alignments)
3609.757 Million cell updates/sec

Title: US-09-724-126a-2
Perfect score: 9224
Sequence: 1 MADEPAGGTERMEISAEPLQ.....ETIARQETNOMLFGFNWQLL 1749

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_ll01.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8617	93.4	1757	20 AAW84351	Murine ubiquitin-p
2	8617	93.4	1757	22 AAB31162	Amino acid sequenc
3	4209	45.6	811	22 AAB93464	Human protein sequ
4	2710.5	29.4	1109	20 AAY02376	Polypeptide identi
5	1716	18.6	333	20 AAW84353	Partial human ubiq
6	1716	18.6	333	22 AAB31163	Amino acid sequenc
7	296	3.2	250	22 AAM25572	Human protein sequ
8	295	3.2	247	21 AAB38334	Human secreted pro
9	197.5	2.1	2096	21 AAB41592	Human ORFX ORF1356
10	189.5	2.1	2048	22 AAM40027	Human polypeptide
11	184	2.0	1411	17 AAW02258	Nucleolar/endosoma

12	182	2.0	2442	21	AA77575	Human cytoskeletal
13	176	1.9	2688	22	AAW40883	Human polypeptide
14	175.5	1.9	1427	12	AAR10534	Human 160kD mediat
15	175.5	1.9	2663	22	AAW39097	Human polypeptide
16	174.5	1.9	1392	20	AAW09999	Restin protein seq
17	173	1.9	2301	22	AAU14720	Novel bone marrow
18	172.5	1.9	1780	22	AAW38681	Human polypeptide
19	172	1.9	1788	22	AAW40467	Human polypeptide
20	170	1.8	2298	22	AAU14794	Novel bone marrow
21	164.5	1.8	2139	22	ABA47278	PN7771 Homo sapi
22	164	1.8	2954	20	AAW01632	Amino acid sequenc
23	156.5	1.7	2710	17	AAW95016	C. difficile toxin
24	156.5	1.7	2710	19	AAW68387	Clostridium diffic
25	155	1.7	3685	10	AAW90290	Human Duchenne mus
26	154	1.7	1562	21	AAW39233	Arabidopsis thalia
27	154	1.7	3685	10	AAW90373	Sequence encoded b
28	152.5	1.7	980	21	AAW18294	Plasmodium faicipa
29	152.5	1.7	1325	18	AAW19540	Male-enhanced anti
30	152.5	1.7	1325	20	AAW94391	Mouse male enhance
31	151	1.6	1521	21	AAW39235	Arabidopsis thalia
32	151	1.6	1528	21	AAW39234	Arabidopsis thalia
33	151	1.6	1703	21	AAW36714	Arabidopsis thalia
34	151	1.6	1710	21	AAW36713	Arabidopsis thalia
35	151	1.6	1744	21	AAW36712	Arabidopsis thalia
36	151	1.6	5447	22	AAU14697	Novel bone marrow
37	150	1.6	5373	22	AAU14603	Novel bone marrow
38	149	1.6	3433	18	AAW22017	Utrophin. Homo sa
39	148	1.6	3021	18	AAW19661	ATM mutant 6348del
40	145.5	1.6	677	20	AAW85608	Secreted protein c
41	145.5	1.6	677	22	AAW64560	Human myosin heavy
42	144	1.6	3003	18	AAW19658	ATM mutant 7630del
43	143.5	1.6	3262	21	AAW53675	Mechanical stress
44	143	1.6	1150	22	AAU14626	Novel bone marrow
45	142.5	1.5	1398	21	AAW18292	Plasmodium faicipa

ALIGNMENTS

RESULT 1
AAW84351
ID AAW84351 standard; Protein; 1757 AA.
AC AAW84351;
XX
XX 25-MAR-1999 (first entry)
XX Murine ubiquitin-protein ligase Ubrl.
XX Ubiquitin-protein ligase; Ubrl; mouse; ubiquitinylation; degradation;
XX N-end rule pathway; stress-related muscle wasting; inhibitor; screen.
XX Mus sp.
XX US5861312-A.
XX 19-JAN-1999.
XX 02-DEC-1997; 97US-0982956.
XX 02-DEC-1997; 97US-0982956.
XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX Kwon YT, Varshavsky A;
XX WPI: 1999-130395/11.
XX N-PSDB: AAW99308.
XX Mouse and human Ubrl cDNA - useful for producing recombinant Ubrl
XX polypeptides
XX Disclosure; Columns 15-28; 18pp; English.
PS

XX The present sequence represents a ubiquitin-protein ligase called Ubr1.
 CC The Ubr1 enzymes are involved in protein ubiquitinylation and
 CC ultimate degradation through the N-end rule pathway and have been
 CC linked to stress-related muscle wasting. Recombinant Ubr1 polypeptides
 CC can be used to screen for inhibitors of muscle wasting when this is
 CC associated with the N-end rule pathway.
 XX
 SQ Sequence 1757 AA;

Query Match 93.4%; Score 861.7; DB 20; Length 1757;
 Best Local Similarity 92.1%, Pred. No. 0;
 Matches 1619; Conservative 64; Mismatches 66; Indels 8; Gaps 3;

QY 1 MADEEAGGTMEIEISAEPLQTPQRLASWMDQVDFYAFHLHLAQLVPEIYFAEMDPDLE 60
 DB 1 madeemdaermdvpeplapqrpsawdqvdqyftafhlhlaqlvpeiyfaemdpdle 60

QY 61 KOESVQMSIFTPLEWYLEGEDPDICLEKLKHSAGFQLCGRVFKSGETYSRCDAIDPT 120
 DB 61 kquesvqmsiltplewylfgedpdicleklkhsagfqlcgkvfksgettyrcrdaidpt 120

QY 121 CVLWDCFDOSVHKHRYKWHSTGCGFCDCGDTFAWKTPFCVNHPEPGRAGTIKENSRC 180
 DB 121 cvlwmcdcfdosvkhkrykwhstggfcgdcgdtfawktpfcvnhpepgragtikeslhc 180

QY 181 PLNEEVIVQARKIFPSVIKYVYEMTIWEPEKELPPBLOIIRKERNERYCVLFNDEHSHSYDH 240
 DB 181 plneeviaqarrifpsvkiyvemytiweekeelpbelpelqirekneryyvclfndehshsydh 240

QY 241 VYISLQALDCELAELAAQLHTTAIDKEGRRAVKAAGYAAQCEAKEDTKSHSENVVQHPHV 300
 DB 241 vyiqlalldcelaelaqlhttaidkegrravkagvyatqceakedikshsnvshpvhv 300

QY 301 EYVLSHSEIMAHOKFALRGLSMWKNKINSYSSDFRQIFCOACLREPDSENCILSLRLMDA 360
 DB 301 evlshseimahokfalrglsmwknkimsyssdfrrqifcoaclreepdsencilslrlmda 360

QY 361 KLYKGARKILHELIFSSPFMEYKFLFAMEFYKYYKQLOKQEYISDDHRSISITALSVO 420
 DB 361 klykgarkilhelifsspfmeneykklfamefykyykqlqkeyisddhersiitalsvq 420

QY 421 MFTVPTPLARHLEEQNVISVIFETLEVLPEYLDNRNKNFPGYSDOKLGRVYAVTCDLK 480
 DB 421 mltvptlarhleeqnvisevitellvpeyldrnknfnfgysqdklgrvyavtcldk 480

QY 481 YILISKPTTWTBRLMQFLEGRFSLKILTCMQGHEIIRQVGOHLEVPDPEAAIAQM 540
 DB 481 yiliskpvtwtrlaqlfgrfslkiltcmqgmeeirrvqghievdpdweaaiaqm 540

QY 541 QLKNTILLFQEWCADEELLVAYKECHKAVMRCSTSFISSTKVQSCGHSLETKSYRV 600
 DB 541 qlkntillfqewcadedllvaykechkavmrctsfisstkvqscghsletksyrv 600

QY 601 SEDLSVHLPLSRTLAGHLVRLSRLGAVSRLEHFEVSFEDQVEVLVEYPLRCLVLVAQV 660
 DB 601 sedlsvhlplsrtaghlvrlsrlgavsrlehefvsfedyvevlveypplrclvlvaqv 660

QY 661 AEMWRRLGSLISQVIFYODVCREMYDKDIIMLQIGASLMDPNKFLLLVLQRYELAE 720
 DB 661 aemwrrlgsllsqvifyyqdvcreemydkdiimlqigaslmdpnkflllvlqryelta 720

QY 721 FNKTTSTKQDQDLIKOYNTLIEMLQVLIIVGERYVPGVGNVTKBEVTMREITHLLCIEP 780
 DB 721 fnkttstkddqdlkoyntlieemlqvliivgeryvpgvgnvtnkbevtmreithllciep 780

QY 781 MPHSATAKNLPENENNETGLENVINKVATFKPKPGVSGHGVYELKDESLKDFNNMYFYHSK 840
 DB 781 mphsaiaknlpennennetglenvinkvatfkkpgvsgghgvylkdeslkdfnmfyfhs 840

QY 841 TQHSKAEHMQKRRKQENKDEALPPPPPPFCFAFSKVINLLNCIDIMYILRTVPERAID 900
 DB 841 tqhskaehmqkrrkqenkdalppppppfcfafsksvinllncidimylrtvperaid 900

DB 841 tqhskaehmqkrrkqenkdalppppppfcfafsksvinllscdmvillrtiiferavd 900

QY 901 TDSNLWTEGMLQMAFHILALGELLEKQOLQKAPEEVEVDFYHKASRLGSSAM--NIQM 957
 DB 901 tesnlwt egmlqmafhilal gellekqolqkapeeevafdfyhkasrlgssamna qnm 960

QY 958 LLEKLKGIPOLEGOKDMITWILQMPDVTVKRLREKSKLIVATTSGSESTKNDLTHDKERA 1017
 DB 958 llerklkgipolegokdmitwilqmpdvtvkr lreksclvattsgleciseethdkeka 1020

QY 1018 ERKRKAAARLHROKIMAOVSALOKNFETHTKMLDNTSEMPGKEDSIMEESTPAVS DY 1077
 DB 1021 erkrkcaarlhrokimaqmsalqknfiethklmlydnsevtgkedsimeest savsea 1080

QY 1078 SRIALGKRGKPSVTEKEVILTCILCOEQEVKTIENAMVLSACVQKSTALTQHRGKPIELS 1137
 DB 1081 srialgkrgpavtekeviltcilcoeqevklennamvlsacvqksta ltrgkpvdlh 1140

QY 1138 GEALDPLFMDPDLAYGYTGTSCGHVHMAVCWQKYFEAVOLSSQQRHVDLDFLESGEYLC 1197
 DB 1141 getldplfmdpdlaygtgtscghvmhvacwqkyfeavqlssqqrhvd ldflesgeylc 1200

QY 1198 PLCKSLCNTVPIIIPLOPKINSENADALAOILLIARWITQTVLARI SGYNIRHAKGENP- 1256
 DB 1201 plckslcntvpiiplopkinsenaealqlltiarwigtvlarisgynirhakgeapa 1260

QY 1257 IPIFFNMGDSTLEFFHSILSGFVSESSIKYSNIKEMVILFATTIYRIGLVKVPDPDR 1316
 DB 1261 ipvifngmgdstlefhsilsgfvgssvksnikemvillfattiyriglvkvpdpdldr 1320

QY 1317 VPMLTWSTCAPTIQAIENLLGDEGKPLFGALQNRHNGLKALMQFAVAQRITCPQVLIQK 1376
 DB 1321 vpmmtwtcatfiqaienllgdegkplfgalqnrhnglkalmqfavaqratcpqvlhik 1380

QY 1377 HLVLISLVLPNIRKSDPTCLLSIDLPHVLGAVLAFPSLYWDDPVDLQPPSSVSSYNHL 1436
 DB 1381 hlarlsvilpnirsdptcpgllsvdlphvlvgavlafpslywddcvdlqpsplssynhl 1440

QY 1437 YLFHLITMAHMLQILLTVDT---GLPLAQVQEDSEEAHSASSFFAEISQYTSSTGCDI 1492
 DB 1441 ylfhlitmahmlqillttdtldspgplaegeedseearcasaffvevshqtdgltcgca 1500

QY 1493 PGWYLVSLKNGITPYLRCAALFPHYLLGVTPPELHTNSAEGEYSALCSYLSLPTNLP 1552
 DB 1501 pgwylvslsrlngitpylrcaallfhylligvappeelfansaegetsalcsylslptnlfl 1560

QY 1553 LFOEYWDTVREPLQRCWADPALLNCLCKQKNTVVRYPKRNSLIETLDDYSCLLNOASHPR 1612
 DB 1561 lfgeywdttriplrgwcdpallkslkqksavrvyrprknslietlpedyscllnqashfr 1620

QY 1613 CPRSADDERKHPVLCLFCGAILCSNICQCEIVNGEEVGACITFHALHCGAGVCIPLKIRE 1672
 DB 1621 cprsadderkhvplclfcgailcsqnicccqivngeevgacvfhalhcgagvciflkire 1680

QY 1673 CRVVLVECKARGCAYPAPYLDYGETDPGLKRGPNPLHLRSERYRKLHLVWQOHCITEETA 1732
 DB 1681 crvvlvckargcaypapyldygetdpglkrpnplhlrseryrklhlv wqohciieeta 1740

QY 1733 RSQETNOMLFGFNWQLL 1749
 DB 1741 rsqetnmlfgfnwql 1757

RESULT 2
 AAB31162
 ID AAB31162 standard; Protein; 1757 AA.
 XX
 AC AAB31162;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Amino acid sequence of Mouse Ubr1 protein.
 XX

Ubr1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
Yersinia enterocolitica; muscle wasting; infection.

Mus sp.
US6159732-A.
12-DEC-2000. 99US-0228317.
11-JAN-1999;
02-DEC-1997; 97US-0982956.
(CALY) CALIFORNIA INST OF TECHNOLOGY.
Kwon YT, Varshavsky A;
WPI; 2001-090278/10.
DR N-PSDB; AAC86933.

Inhibiting the N-end rule pathway in mammalian cells for treating
infections and various diseases associated with muscle tissue wasting,
by inhibiting the expression of Ubr1 gene

Example; Column 15-28; 18pp; English.

The present sequence represents a murine Ubr1 enzyme. Ubr1 is an E3-type
protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
ligase. The enzyme is specific for destabilising residues exposed at
the N-terminus of protein substrates. Inhibition of the expression of
Ubr1 gene in a cell results in inhibition of the N-end rule pathway.
The method is used for treatment of mammalian cells infected with an
intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
enterocolitica. Inhibition of N-end rule pathway is also useful for
treating various diseases associated with wasting of muscle tissue and
infections.

Sequence 1757 AA;

Query Match 93.4%; Score 8617; DB 22; Length 1757;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1619; Conservative 64; Mismatches 66; Indels 8; Gaps 3;

QY 1 MADEEAGTERMETSAELPQTPQRLASWDDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 60
DB 1 madeemgaermvdspepplapgrpaswddqvdftytaflhhlagvlpeiyfaemdpdl 60

QY 61 KOESVQMSITPLEWYLFQGDPOICLEKLGKSHGAFOLCGRVFKSGETYSRCDCAIDPT 120
DB 61 keesvqmsitplewylfqedpdaiclekshgafqlcgkfvksgetyscrdcaidpt 120

QY 121 CVLMDCFQDSVHKHRYKMHSTGTGGGFCDCGDTFAWKTPFCVNHPEPRAGTIKENSRC 180
DB 121 cvlmdcfqsvhkhryknhstgtggfcdcgdtewktpfcvndhepragttikeslhc 180

QY 181 PLNEEIVQARKFPSPVIVKVVEMTIEEEKELPPELQIREKNERYCVLENDEHSHSDH 240
DB 181 plneevlaqarrifpsvikviventieeekelppeiqireknerycvlendehshsydh 240

QY 241 VIYSLQALDCELAELAAOLHTTAIDKEGRRAVKAGAYAACQEAKEEDIKSHSENSQHPHVV 300
DB 241 viyslqalcdelaqaqlhttaldegravagvayacqeaikedikshsensvqhphlv 300

QY 301 EVLSEIMAHQKFAIRLGRSWMNKIMSYSSDFQIFCQACLRPEEPDSENPCLISRLMLWDA 360
DB 301 evlsvmahqkfairlgrswmnkimsyssdfqifcqaclreepdsepnpcclisrlmlwda 360

QY 361 KLYGARKILHELIFSSFFMEMEYKKLFAMFVYKYQLOKQKEYISDDHDSRISITALSVQ 420
DB 361 klygarkilheliffssffmemeykklfamefvkykqlqkeyisddhdersitalsvq 420

QY 421 MFTVPTLARHLIEEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGKRVYAVICDLK 480
DB 421 mltvptlarhlieeqnvvisvitetllevlpeylldnrnknfqqysdqklgkryvavicdlk 480

QY 481 YLISKPTTWTERLRMQFLGFRSFLKILTCMQGMEIRQVQGHIEVDPPWEAIAIQM 540
DB 481 ylisxpvtwterlraqflgfrsfkiltcmqgmeeirrvqvgqhievdppweaiaiqm 540

QY 541 OLKNILLMFQEWACDEELLVAYKECHKAVMRCSTSFISSTKVTVOSCGHSLSTKSYRV 600
DB 541 qlknillmfqewacdedlllvaykechkavmrcstsfisstkvvtvoscghslstksyrv 600

QY 601 SEDLSIHPLSRTLAGLHVRLSRLGAVSRLEHFVSFEDFQVEVLVPEYRCLVLVAQVW 660
DB 601 sedlsvihplsrtaglhvrlsrलगaisrlhefvpfidsqvevlvpeyrlcvlvaqv 660

QY 661 AEMWRRNGLSLSQVFFYYQDVVKCREMYDKDIIMLQITGASLMDPNKFLLLVLQRYEAE 720
DB 661 aemwrrnglslsqvfyvqdvkcreemydkdiimlqigaslmdpnkflllvlqryeae 720

QY 721 ENKTISTKDDLLKQYNTLIEEMLOVLIVGERYVPGVGNVTKKEEVTMRFIHLICIEP 780
DB 721 fntkistkddllkqyntlieemlvllivgeryvpgvgnvtkreemvtrfihliciep 780

QY 781 MPHSAIAKNLPENNNPTGLENVINKVATPKPCVSGHGVYELKDESLKDFNMFYFYSK 840
DB 781 mphsaiaknlpennnetglenvinkvatfkpgvsgbhvyeikdeslkdfnmfyfysk 840

QY 841 TOHSAEAMOKRRKQENKDEALPPPPPEFCFPAFSKVINLLNCDIMMYLRTVFERAID 900
DB 841 tqhsaehmkrrkqenkdalpppppefcfpafskvinnllncdimmylrtlrtvferaid 900

QY 901 TDSNLWTEGMLQAFHILALGILLEEKQLOKAPEEYTFDFYHKASRLGSSAM---NQM 957
DB 901 tesnlwtegmlqafhialalglleekqlkapeeevafdfyhkarslgsammaqnlm 960

QY 958 LLEKLKGIQLEGQKQDMITWILQMFDTVKRLREKSKLIVATTSSSESTKNDIEITHDEKA 1017
DB 961 llerlkgiqlegqkdmiltwilmfdtkvrlreksklivattsssestkeieithdeka 1020

QY 1018 ERKRKAARLHRQKTIQAQNSALQKNFIETHKLMYDNTSEMPGKEDSIMEEESTPAVS DY 1077
DB 1021 erkrkaearlhrqkimaqnsalqknfiethklydntsevtgkedsimeeestsavsea 1080

QY 1078 SRLALGPKGPSYTEKEVLTCLCOBEOEYK IENNAVLNACVOKSTALQHRCKPIELS 1137
DB 1081 srlalgpkrgpavtekevltclcoqeevklennamvlsacvdkstaltqhrckpvdhl 1140

QY 1138 GEAIDPLFMDPLAYGTYTGSCGHVMHVCWKYFEAVQLSSQORIHVDLFDLESGEYLC 1197
DB 1141 getldplfmdplahytytgscghvmhvcwkyfeavqlssqqrhvdldflesgeylc 1200

QY 1198 PLCKSLCNTVPIPILOPKINSADALAOQLTLARMIQTVLARISGYNIRHAKGNP- 1256
DB 1201 plcksicntvpiplpqpkinsenacalaqltlarwiqtvlariisgynikhakgeapa 1260

QY 1257 IPIFFNQMGDSTLEFHSILSFGVSESIKYSNISKEMVILPATIYRIGLKVPPDERDPR 1316
DB 1261 vplifnqmgdstlefhsilsgvsvksysikemvilfattyiriglkvpvppdeidpr 1320

QY 1317 VPMLTWSTCAFTTQATENLLGDEGKPLFGALQNRHQNGLKALMOFAVAORITCPOVLIOK 1376
DB 1321 vpmmtwtstcafttqatienllgdegkplfgalqnrhqsnglkalmqfavaqratcpqvlilh 1380

QY 1377 HLVRLLSVVLPNIKSEDTPCLLSIDLFLHVLGVAVLAPSPSYWDDPVPDPSSVSSSYNHL 1436
DB 1381 hlarllsvlvpnlqsentpgllsvdlfhvlvgavlaflpslywddtvdqlqpsplssynhl 1440

QY 1437 YLFHLITMAHMLQILLTVDT- ---GLPLAQVQEDSESAHSASSFSAEISQVTSIGCDI 1492
DB 1441 ylfhlitmahmlqillttdtldspgpplaegeedseearcasatfvevsgtdlgtcga 1500

QY 1493 PGWYLVSLKNGITPYLRCAALFFHLLGYTPPELHTNSAEGEYSALCSYLSLPTNLFL 1552

RESULT 4	
AA02376	
ID	AA02376 standard; Protein; 1109 AA.
XX	
AC	AA02376;
XX	
DT	13-JUL-1999 (first entry)
XX	
DE	Polypeptide identified by the signal sequence trap method.
XX	
KW	Signal sequence trap method; SST method; immunisation; inhibition;
KW	infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW	activin activity; inhibin activity; chemokine activity;
KW	cytokine activity; blood coagulation regulation; agonist; antagonist;
KW	metabolic disorder; hormonal disorder; immune disorder; wound;
KW	severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer.
XX	
OS	Homo sapiens.
XX	
PN	WO9918126-A1.
XX	
PD	15-APR-1999.
XX	
PF	06-OCT-1998; 98WO-JP04514.
XX	
PR	07-OCT-1997; 97JP-0274674.
XX	
PA	(ONOY) ONO PHARM CO LTD.
XX	
PI	Fukushima D, Shibayama S, Tada H;
XX	
DR	WPI; 1999-277254/23.
DR	N-PSDB; AAX35730, AAX35731.
XX	
PT	Polypeptides identified by the signal sequence trap method from a
PT	human cDNA library
XX	
PS	Claim 1; Page 162-167; 281pp; Japanese.
XX	
CC	AA02358-84 represent novel polypeptides which are identified from a
CC	human placental cDNA library by the signal sequence trap (SST) method.
CC	The polypeptides are encoded by the cDNA sequences in AAX35694-X35747.
CC	The polypeptides have a broad range of physiological activity, including
CC	immunisation against and inhibition of infections, allergies and cancer;
CC	regulation of tissue formation and repair; activin/inhibin activity;
CC	chemokine/cytokine activity; blood coagulation regulation; and
CC	receptor/ligand agonist or antagonist activity. The polypeptides can
CC	be used for prevention and treatment of disorders including infections
CC	by bacteria, yeasts and viruses (including HIV) and protozoa; metabolic
CC	and hormonal disorders; immune disorders (including severe combined
CC	immunodeficiency (SCID) and AIDS; thrombosis; cancer; and traumatic or
CC	surgical wounds.
XX	
SQ	Sequence 1109 AA;
Query Match	29.4%; Score 2710.5; DB 20; Length 1109;
Best Local Similarity	46.9%; Pred. No. 3.4e-220;
Matches 529; Conservations 214; Mismatches 345; Indels 39; Gaps 18;	
Qy	644 VLVEPLRLVLAQVVAEMWRNGLSLISQVYQDVKCREEMDKDIIIMLGASIMD 703
Db	1 mliehpirlclvcaqvhagmrrngfslvnuqlyyhnkrrcmfkdavmqlgtgvmmd 60
Qy	704 PNKFLLVQLQRYELAEAFN-----KTISTK--DQDLIKQYNTLIEEMLQVLIYIVGERYV 756
Db	61 pnhflmlsrfrlyqlfstpdvgkrfsseithkdvvgqntlieemlyliimlvgerfs 120
Qy	757 PCGVNWTKEVWREIRIHLLCIEPMPHISAIKNNPENNETGLENIYINKVATFKKPGVS 816
Db	121 pvgvgnatdeikreihlqslkpmahselvkslpedenketgmvesieavahfkpglt 180
Qy	817 GHGVVELKDESKDFNMVYHYHYSKTHQSKAEHMQKKRRKQENKDEALPPPPPPFCFPAFS 876
Db	181 grgmyelkpecakefnlyfhrsraeqskaeagrkikrnpredtalppvpfpfcplfa 240
Qy	877 KVINLLNCIDIMMYILRTVPERAIDTDSNLNTEGMLQMAFHILALGLEEKQOLKAPREE 936
Db	241 slvnllqsdvmlcimgtliqwavehngyawsesmlqrvlhlgnalgeekqhlenvtee 300
Qy	937 -VTFDFYHKASRLGSSAMN---IQMLLEKLGIPOLEGGQDMITWILQIPMTFVKRLREKS 992
Db	301 vvtftftqiskpgeapknspisilamletlqnapylevhkdmirwilktfnvavkkmress 360
Qy	993 -CLIVATTSGSEIKNDEITHDKAEARKKAEARLHRKIMQAQMSALQKNFETFKLM 1051
Db	361 pcpvaetegt1m--eessrdkkaerkkaelarlrrkimaqsemqrhidenkel 417
Qy	1052 YNTSMPEKESIMEEESTPAVSDYSRIALGPKRGSPVTEKEVLTCLCOEEOEVKTEN 1111
Db	418 fqgtleldastavldh--spvasdmtlaltalgaqtqvpegrqvtvtcllceeqevkves 475
Qy	1112 NAMVLSACVOKSTALTQHRGKPIELSGEALDPLFMDPLAYGTYTSGGHVMHACVQKY 1171
Db	476 ramvlaafvgrstvlsknrskfiq-dpekydplfmhpdlsctgtstscghlmahcwary 534
Qy	1172 FEAVQLSSQO-----RIHVDLFDLESGEYLCPLCKSLCNTVPIPIQPOKINSENADAL 1226
Db	535 fdsvqakeqrqrllrlhts-ydvengeficpiceclstvtipil-lsprnfnrnln-f 591
Qy	1227 AQLLTARWQTQVLARISGYNIRHAKGENPIPIFFNMGMDSTLEFHSISFGVSESSIKY 1286
Db	592 sdqpnltqwtitsgqikalqfrrkeestpnastknsenvdelqlpegfrpdrprkipy 651
Qy	1287 SNSIREMVLFTATTIYRIGLKVPPDERDPVPMLTWTSCAFTQAIENLLCDEGKPLFGA 1346
Db	652 sesikemltftgtatkykvlgvhpnneedprvpmcwscaytisqiserilsdedkplfpg 711
Qy	1347 LQNRQHNGLKALMQFAVAQRITCPQVLIQKHLVRLSWLPNIKSEDTPCLLSLDLFHV 1406
Db	712 lpcriddeclrsitrfaaahwtvasvvgqhfcklfaslvondsheelpcildidmflhl 771
Qy	1407 VGAVLAPPSLYDDPVDLQPPSSVSSYNHLYLFHLITMAHMLQILL---TVDTGGLPLAQV 1463
Db	772 vglvlatfapalqcd-----fsgislgtdghlhlvtmahliqilltscsteengmd--qe 824
Qy	1464 QEDSEAHASAFFAISQYTSIGSICDIP-GWLVWVSLNGITPYLRCAALFHYLLGV 1522
Db	825 nppeceesavlaalyktlhyt-gsalkeipsghlwtvsvragimpfikcsalfhylnvg 883
Qy	1523 TPPEELHTNSAEGEYSALCSYLSLPTNLFLLFQBYWDTVRPLQWQCADPALLCCKOKN 1582
Db	884 psppdiqv-pgtsfhfehlesylslpnnliclfqenseimmslieswornsevkryleger 942
Qy	1583 TVVYPRKRNSLIELEDDYSCLLNQASHFRCPRSADDERHPVLCFLFCGALICSONICQ 1642
Db	943 dairypresnklinlpedyslinqasnfscpksggdkstaptclvcgslcsqsyccq 1002
Qy	1643 ETVNGEEVACAFHALHCGACVCIFLKIRECRVVVLVBGKARGCAYPAPYLDYGETDPGL 1702
Db	1003 telegedvgactahtyscsgvgvflrvrcqvflagktkgcfysppylidygetddgl 1062
Qy	1703 KRGNPLHLSRERYKLUHVMQOHCITEEIARSOFTNOMLFGFNQOLL 1749
Db	1063 rrgnplhlckerfkqklwhqhsvteeighadeanqlvgidwqhl 1109
RESULT 5	
AAW84353	
ID	AAW84353 standard; Protein; 333 AA.
XX	
AC	AAW84353;
XX	
DT	25-MAR-1999 (first entry)
XX	

DE Partial human ubiquitin-protein ligase, Ub1.

XX Ubiquitin-protein ligase; Ub1; human; ubiquitinylation; degradation;
 KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen.
 XX Homo sapiens.
 XX US5861312-A.
 XX 19-JAN-1999.
 XX 02-DEC-1997; 97US-0982956.
 XX 02-DEC-1997; 97US-0982956.
 XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
 PA Kwon YT, Varshavsky A;
 PI WPI; 1999-130395/11.
 DR N-PSDB; AAX03300.
 XX Mouse and human Ub1 cDNA - useful for producing recombinant Ub1
 PT polypeptides
 XX Disclosure; Columns 27-30; 18pp; English.
 XX The present sequence represents a partial ubiquitin-protein ligase called
 CC Ub1. The Ub1 enzymes are involved in protein ubiquitinylation and
 CC ultimate degradation through the N-end rule pathway and have been
 CC linked to stress-related muscle wasting. Recombinant Ub1 polypeptides
 CC can be used to screen for inhibitors of muscle wasting when this is
 CC associated with the N-end rule pathway.
 XX Sequence 333 AA;

Query Match 18.6%; Score 1716; DB 20; Length 333;
 Best Local Similarity 99.7%; Pred. No. 9.9e-137;
 Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 702 MDPNFKLLVLQRYELAEAFNKTISTKDDLIKQYNTLIEEMLQVLIIYIGERYVPGVGN 761
 DB 1 mdpnkfllvlqryelaeafnktistkdqlikqyntlieemlqvliiyigeryvpgvgn 60

QY 762 VTKEEVTMRIIHLCTIEPMPSAIAKNLPENENNETGLENVINKVATFKKPGVSGHGY 821
 DB 61 vtkeevtmreiihlctiepmphsaiaaknlpennennetglenvinkvatfkkpgvsgngvy 120

QY 822 ELKDESLKDFNMVYHYSKTQHSKAEHMKRRKQENKDEALPPPPPEFCFAFSKVINL 881
 DB 121 elkdeslkdfnmvyfhyssktqhskaehmqkrrkqenkdealpppppfcfapfskvinl 180

QY 882 LNCDDIMMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKKQLOKAPEEVTFDF 941
 DB 181 lncddimmyilrtvferaintdsnlwtegmqlmafihailalglleekqqlkapeeevtfdf 240

QY 942 YHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWILQMDTVKRLREKSKLIVATTSG 1001
 DB 241 yhkasrlgssamniqlmleklgipqllegqkdmittwilqmdftvkrlekreksclivattsg 300

QY 1002 SESIKNDEITHDKAEKRRKAEAAARLHRQKIM 1034
 DB 301 sesikndeithdkeakerrkaeaaarlhrqkim 333

RESULT 6
 AAB31163
 ID AAB31163 standard; Protein; 333 AA.
 XX AAB31163;
 AC AAB31163;
 XX 02-APR-2001 (first entry)

XX Amino acid sequence of a partial human Ub1 protein.
 DE Ub1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
 KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
 XX Yersinia enterocolitica; muscle wasting; infection.
 XX Homo sapiens.
 XX US6159732-A.
 XX 12-DEC-2000.
 XX 11-JAN-1999; 99US-0228317.
 XX 02-DEC-1997; 97US-0982956.
 XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
 PA Kwon YT, Varshavsky A;
 PI WPI; 2001-090278/10.
 DR N-PSDB; AAC86934.
 XX Inhibiting the N-end rule pathway in mammalian cells for treating
 PT infections and various diseases associated with muscle tissue wasting,
 PT by inhibiting the expression of Ub1 gene
 XX Example; Column 27-30; 18pp; English.
 XX The present sequence represents a partial Ub1 enzyme. Ub1 is an E3-type
 CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
 CC ligase. The enzyme is specific for destabilising residues exposed at
 CC the N-terminus of protein substrates. Inhibition of the expression of
 CC Ub1 gene in a cell results in inhibition of the N-end rule pathway.
 CC The method is used for treatment of mammalian cells infected with an
 CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
 CC enterocolitica. Inhibition of N-end rule pathway is also useful for
 CC treating various diseases associated with wasting of muscle tissue and
 CC infections.
 XX Sequence 333 AA;

Query Match 18.6%; Score 1716; DB 22; Length 333;
 Best Local Similarity 99.7%; Pred. No. 9.9e-137;
 Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 702 MDPNFKLLVLQRYELAEAFNKTISTKDDLIKQYNTLIEEMLQVLIIYIGERYVPGVGN 761
 DB 1 mdpnkfllvlqryelaeafnktistkdqlikqyntlieemlqvliiyigeryvpgvgn 60

QY 762 VTKEEVTMRIIHLCTIEPMPSAIAKNLPENENNETGLENVINKVATFKKPGVSGHGY 821
 DB 61 vtkeevtmreiihlctiepmphsaiaaknlpennennetglenvinkvatfkkpgvsgngvy 120

QY 822 ELKDESLKDFNMVYHYSKTQHSKAEHMKRRKQENKDEALPPPPPEFCFAFSKVINL 881
 DB 121 elkdeslkdfnmvyfhyssktqhskaehmqkrrkqenkdealpppppfcfapfskvinl 180

QY 882 LNCDDIMMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKKQLOKAPEEVTFDF 941
 DB 181 lncddimmyilrtvferaintdsnlwtegmqlmafihailalglleekqqlkapeeevtfdf 240

QY 942 YHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWILQMDTVKRLREKSKLIVATTSG 1001
 DB 241 yhkasrlgssamniqlmleklgipqllegqkdmittwilqmdftvkrlekreksclivattsg 300

QY 1002 SESIKNDEITHDKAEKRRKAEAAARLHRQKIM 1034
 DB 301 sesikndeithdkeakerrkaeaaarlhrqkim 333

```
RESULT 7
ID   AAM25572 standard; Protein; 250 AA.
XX   AC
XX   AAM25572;
XX   DT 16-OCT-2001 (first entry)
XX   DE
XX   DE Human protein sequence SEQ ID NO:1087.
XX   KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX   KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX   KW antibacterial; endocrine; cardiac; central nervous system; virucide;
XX   KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
XX   KW antiagregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
XX   KW dermatologic; anti allergic; antiasthmatic; antidiabetic; cytostatic;
XX   KW neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection;
XX   KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX   KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX   KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
XX   KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX   KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
XX   KW allergic rhinitis; diabetes; multiple sclerosis; depression;
XX   KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX   KW neurological disorder.
XX   OS Homo sapiens.
XX   PN WO200153455-A2.
XX   PD 26-JUL-2001.
XX   PF 22-DEC-2000; 2000WO-US35017.
XX   PR 23-DEC-1999; 99US-04711275.
XX   PR 21-JAN-2000; 2000US-0488725.
XX   PR 25-APR-2000; 2000US-0552317.
XX   PA (HYSE-) HYSEQ INC.
XX   PI Tang YT, Liu C, Drmanac RT;
XX   DR WPI; 2001-457603/49.
XX   DR N-PSDB; AAH99513.
XX   PT Isolated human polynucleotides encoding polypeptides, useful for the
XX   PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX   PS Claim 20; Page 220; 1217pp; English.
XX   CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX   CC AAM25963. The proteins can have activities based on the tissues and
XX   CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX   CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
XX   CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
XX   CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;
XX   CC antiulcer; osteopathic; dermatologic; antiallergic; antiasthmatic;
XX   CC antidiabetic; cytostatic; neuroprotective; antidepressant; neurotropic;
XX   CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX   CC encoding them can be used in gene therapy, antisense therapy and vaccine
XX   CC production. The proteins and polynucleotides are useful for screening for
XX   CC agonists or antagonists of a protein and for the treatment and diagnosis
XX   CC of disorders associated with the activity of a protein e.g. inflammation,
XX   CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX   CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX   CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX   CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX   CC osteoporosis, severe combined immunodeficiency, eczema, allergic
XX   CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX   CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX   CC neurological disorders.
XX   SQ Sequence 250 AA;

Query Match 3.2%; Score 296; DB 22; Length 250;
Best Local Similarity 30.1%; Pred. No. 16e-16;
Matches 74; Conservative 46; Mismatches 88; Indels 38; Gaps 10;

QY 1500 SLKNGITPYLRCAALFFHYLLGVTPPELHTNSAEGYSALCSYLS-LPTNLFLLFOEYW 1558
DB 7 alqdfclpfirtsllqhhlf-----edlpsqceeeefsvlascglilpt-----ty 54

QY 1559 DTVRPLLRWCAD---PA--LLN--CLKQKNTVVYRPR-----KRNSLIELPD 1599
DB 55 qtehpfiisacldwpyafdiithwcfelksfsterhaegdkalligeskwkplhllqipe 114

QY 1600 DYSCLLNQASHPRCPRSADDERKHPVLCFLFCGAILGSONICCOFIVNGEEVGACIFHALH 1659
DB 115 nntif-qyyhrktesvctkvpkdpavclvcgftvclkgicckq----qsycecvlshqn 169

QY 1660 CGAGVCIFLKIRECRYVLVEGKARGCAYPAPYLDVEGETDGLKRGNPLHLSRERYRKLH 1719
DB 170 cgagtgifllinasvliirgh-rfclwgsvyldahgeedrdrirgkplyckerykvle 228

QY 1720 LVWQQH 1725
DB 229 qqwish 234

RESULT 8
ID   AAB38334 standard; Protein; 247 AA.
XX   AC
XX   AAB38334;
XX   DT 31-JAN-2001 (first entry)
XX   DE Human secreted protein encoded by gene 14 clone HASAU84.
XX   KW Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
XX   KW cytostatic; cardiac; vasotropic; cerebroprotective; neuroprotective;
XX   KW neurotropic; antibacterial; virucide; fungicide; ophthalmological; human;
XX   KW vulnery; gene therapy; infection; secreted protein.
XX   OS Homo sapiens.
XX   PN WO2000061623-A1.
XX   PD 19-OCT-2000.
XX   PF 06-APR-2000; 2000WO-US08979.
XX   PR 09-APR-1999; 99US-0128693.
XX   PR 26-APR-1999; 99US-0130991.
XX   PA (HUMA-) HUMAN GENOME SCI INC.
XX   PI Ruben SM, Ni J, Komatsoulis CA, Rosen CA, Soppet DR, Shi Y;
XX   PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;
XX   PI Young PE;
XX   DR WPI; 2000-647418/62.
XX   PT New nucleic acid molecules encoding 62 human secreted proteins for
XX   PT diagnosing, preventing, treating or ameliorating medical conditions and
XX   PT used as food additives or preservatives -
XX   PS Claim 11; Page 604-605; 716pp; English.
XX   CC Sequences AAB38331-B38396 represent the amino acid sequences of 62
XX   CC human secreted proteins encoded by the genes AAC69512-C69587. The genes
XX   CC and proteins are useful for preventing, ameliorating or treating medical
XX   CC conditions, e.g. by protein or gene therapy. The genes are isolated from
XX   CC a range of human tissues disclosed in the specification. The nucleic
XX   CC acids, proteins, antibodies and (ant)agonists are useful in the
```

CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
 CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms
 CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac
 CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
 CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
 CC infections caused by bacteria, viruses and fungi; and (h) ocular
 CC disorders e.g. corneal infection. The polypeptides can also be used to
 CC aid wound healing and epithelial cell proliferation, to prevent skin
 CC aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis.
 XX
 SQ Sequence 247 AA;

Query Match 3.2%; Score 295; DB 21; Length 247;
 Best Local Similarity 30.2%; Pred. No. 1.9e-16;
 Matches 74; Conservative 43; Mismatches 90; Indels 38; Gaps 9;

QY 1501 LKNGITPYLRCAALFHYLLGVTPPELHTNSAEGYSALCSYLS-LPTNLFLLFOEYWD 1559
 Db 5 lqdfcplfritsllqhlfg-----edlpscgeeeefsvlasclgilpt-----fyq 52
 QY 1560 TVRPLLORWCAD--PAL---LNCILKOKNTVVRYPR-----KRNLSLIELPDD 1600
 Db 53 tehpffisaslcpwpvpafdlitqwcfeiksftrhaeqgkalligskwklphllqipen 112
 QY 1601 YSCLLNQASHFRCPRSADDERKHPVLCFGALCSQNICCCQIVNGEEVGACIFHALHC 1660
 Db 113 yntif-qyyhrktsvctkvpdpavclvcgtfclkgicckq---qsycecvlshqnc 167
 QY 1661 GAGVCIFLKRRCRVVLVGKARGCAYPAPYLDYGETDPLGRGNPLHLRSRYRKHL 1720
 Db- 168 gagtgifllinasviiiirgh-rfcilwsgvyldahgeedrldrrgkplyickerykvleg 226
 QY 1721 VVQQH 1725
 Db 227 qwish 231

RESULT 9
 AAB41592
 ID AAB41592 standard; Protein; 2096 AA.

XX AAB41592;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF1356 polypeptide sequence SEQ ID NO:2712.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipruritic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disease; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX

PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC75801.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 1952-1956; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipruritic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 2096 AA;

Query Match 2.1%; Score 197.5; DB 21; Length 2096;
 Best Local Similarity 17.3%; Pred. No. 1.3e-06;
 Matches 259; Conservative 236; Mismatches 518; Indels 485; Gaps 60;

QY 117 IDPTCVLCMDCFQDSVHKHRYKMTSTGGCGFCGCDTEAMKTFGFCVNHPEGRAGTKE 176
 Db 645 ldetvsvckka-qenmkqrhenethl-----eeqtsdlkmkkaelqggaavike 693
 QY 177 ---NSRCPLNEEVIVQARKIFPSVIKYVEMTWIWEKEELPPELOIREKNERYCYVLFND 233
 Db 694 ahheatcrheee-----kkqlqvkleekthlqekrlqcheme-----731
 QY 234 EHSYDHYVLSQALDCELAQAQHTTAIDKGRRAYKAGY-----AACQEAKEEDIK 287
 Db 732 -----lkarlqag---asfgr-reglqssaateekvrglqleq---769
 QY 288 SHSEN---VSQHPHVEVLHSEMAHOKFALRLGSGWNKIMSYSSDFRQFCQACLEEE 343
 Db 770 fhqeqltslvkehkleekrlqleq-----fek 807
 QY 344 PDSNPCLISRLMDAKLYKGARKILHELIFSSPFMEYKKLFAEFVYKYOLOKEY 403
 Db 808 metecnrrtsqi---eaqfscdqkvtercsalqslqgryr-----qelkqlqeqq 856
 QY 404 ISDDHRSISITALSVMFTVPTLARHLIEQNVTISVITETLLEVLPE-YLDRNNKF--N 460
 Db 857 reeksqwefekdeltqcaeaqellketlkrkettstlvtqeremlektykdhlnsmvve 916
 QY 461 FQGSYQDKLGRVAVICDLKYILISKPTIWTER---LRMQFLEGFRSLKILTCMQGM-F 516

[illegible]

SO	Sequence	2688 AA;
	Query Match	1.9%; Score 176; DB 22; Length 2688;
	Best Local Similarity	18.6%; Pred. No. 0.00013;
	Matches	Conservative 310; Mismatches 265; Indels 524; Gaps 80;
QY	174 IKNSRCPLNEEVIVVQARKTTPSPVIKYVVEMTWEEKEKLPPEL-----OIREKNERYYC 228 : : : : : : : : : :	
DB	757 veene--alreevi-----llselkslpseverlrkeiqgkseel-- 794 : : : : : : : : : :	
QY	229 VLFNDEHHSDHYIYSIORALDCELAFAQLHTTAIDRKGRAVKAGAYACQAKEDI-- 286 : : : : : : : : : :	
DB	795 -----hiits-----ekdkifsevnhkesr---vqglleeigtkddlat 831 : : : : : : : : : :	
QY	287 -KSHSENVSOHPHLVEVLHSEIMAHOKFALARLGSWNN-KIMSYYSSDFROIFCQACLREEP 344 : : : : : : : : : :	
DB	832 tqsnystdqefgnftlmdfegkykmvleenernmqeivnliskeaqkf----- 881 : : : : : : : : : :	
QY	345 DSENPCILISRLMDARLYKGARKI---LHE-----LIFFSF 378 : : : : : : : : : :	
DB	882 dsslgalktelsyktgeltektrevqerlnemeqlkeqlenrdspqtverehtltitek1 941 : : : : : : : : : :	
QY	379 FMEWEYKKLFAMEFYVYKQLOK-----EYTSDDHDR-SISITALSVOMFTVPTLARH 430 : : : : : : : : : :	
DB	942 qgtleevkttlge-kadlkqlesqierdqlksdhdtvnmmidtgqlrnaleslkhq 1000 : : : : : : : : : :	
QY	431 LIEEQNVISVITETLLLEVLPPE--YLDRNNKFNGGYSDOKLGRVYAVICDLKVYLISK-- 486 : : : : : : : : : :	
DB	1001 ----getintlskiseeivrnlhmeentgetdefqqkmvg-----ldkkqdleaant 1050 : : : : : : : : : :	
QY	487 PTWTEFLRMQFLEGRFSFKILTQCMBEIRRVGOHIEVDPDWEAATAIQMOLK--- 543 : : : : : : : : : :	
DB	1051 qtltadvkdnetieqr--kfislqeknelqgm1-----esvlaekeqlktdl 1097 : : : : : : : : : :	
QY	544 --NILMFQEWCADEBELLVA--YECHKAVMRCSSTSFSSSKTVVQSCG-----HSL 593 : : : : : : : : : :	
DB	1098 kenientien----qeelfilgdellkkqevaqaeknhaikkegelrtsctdrilaeeveekl 1153 : : : : : : : : : :	
QY	594 ETKSYRVSE---DLVSIIHPLS-----RTLAGLVLSRLGAVSRLLH 632 : : : : : : : : : :	
DB	1154 keksqqlikekqqlinvqeemseqbkineienklnekltelhemeterieleagkin 1213 : : : : : : : : : :	
QY	633 EFVSFEFOVELVLEYLRCLVLVAOVAMWRNRGLSIISOVFYQDVKCREMYDKOI 692 : : : : : : : : : :	
DB	1214 e--nyee-----vkstke--rkvlkelqsfeterdhlrgyireia 1252 : : : : : : : : : :	
QY	693 IMLIQGASLMDPNKFIILLVORY-ELAEAFNTISTK-----DQDLIKOYNLTLIBEMLQ 745 : : : : : : : : : :	
DB	1253 tgiqtkeel-----kiahihlheqetidelrrsvsektaqiintdgleshtklgee-- 1305 : : : : : : : : : :	
QY	746 VLIVYIGERYVPVGCVNTKEVTMRIIH-----LCIEPMPHSAIAKANLPENE 794 : : : : : : : : : :	
DB	1306 ipwlheeqlpnkvkvetsetcmnelelltqstktdsttlariemerlrinekfqeq 1365 : : : : : : : : : :	
QY	795 NNETGLENVINKATPKPGVGHGVYELKDCLKDNFMFYFH-----YSKTQHASKAEHMQ 850 : : : : : : : : : :	
DB	1366 eeikstkerdnltkike-----alevhhdqlke-----hiretlakigesqsqkeq 1412 : : : : : : : : : :	
QY	851 KRRKOENKDEALPPPPPFPFCFAPFSKVINLLCDIMVIL-----RTV----- 894 : : : : : : : : : :	
DB	1413 slnmkedndntkiwsemegfkpkdsa---llrielemjglskrlqeshdemksvakekd 1469 : : : : : : : : : :	
QY	895 ----FERADTDNSLMTGEMLQ-MAPHIALGLLEEKOOLoka-----PEEEVTDFDYHK 944 : : : : : : : : : :	
DB	1470 dlqrlqviesdsqikenikeivah-----letecekvahcclkeqeteinelrvn 1523 : : : : : : : : : :	
QY	945 ASRLGSANNIOMLLE-----KLKGIPQLEGQKDM---ITWIQMFDTVKRRLREKSCL 994 : : : : : : : : : :	
DB	1524 lseketeistqiqleaindklnkqlkeiyekееqninikisevqevkneltqfkehr-- 1581 : : : : : : : : : :	
QY	995 IVATTSGSESIGND--EITHD-----KEKAERKRKAEARLHR-----QXT 1033 : : : : : : : : : :	

XX WPI: 1991-038913/06.
DR N-PSDB; AAQ10378.
XX
PT 160 kD human polypeptide mediator or precursor of inflammation -
PT polyclonal or monoclonal antibodies to polypeptide treat and
PT diagnose chronic inflammation and hodgkins lymphoma
XX
PS Claim 3; Page 32: 47pp; English.
XX
CC The protein is a cytokine used to treat chronic inflammatory
CC conditions. It is prepared by chromatographically purifying an
CC optionally pre-purified cell extract, cell supernatant or cell
CC filtrate of stimulated normal human leucocytes or human embryonic
CC epithelial lung cells. Alternatively, the protein can be produced
CC by microorganisms or continuous mammalian cell lines, transformed
CC with plasmids encoding MRP-160. The invention also covers the
CC polypeptide fragment from amino acids 878-1427 and derivatives of
CC the protein in which the amino and/or hydroxyl functions are
CC glycosylated or acylated and have mol. wt.s of 190 and 140kD,
CC respectively.
XX
SQ Sequence 1427 AA;

Query Match 1.9%; Score 175.5; DB 12; Length 1427;
Best Local Similarity 19.3%; Pred. No. 5e-05;
Matches 199; Conservative 158; Mismatches 350; Indels 323; Gaps 43;

QY 167 EPCRGATIKENSRCPLNEEVIVQARKIFPSVIVYVEMTWEEKEELPELQ----IREK 222
DB 488 edrvratvseker-----imekdialrvqveaelrrr 521
QY 223 NERYVCVLFNDEHSHYDHYVLSQALDC-----ELAAQLHTTAIDKGRRAVAKAGA 275
DB 522 lesnpgagdvmslsllqeisslqeklevtrtdhgreitslkehfgareethqkeika-l 580
QY 276 YAACQ---EAKEDIKS-----HSENVSQPLHVEVLHSEIMAHQKFPALRLG-SWMNKIMS 326
DB 581 yateklskeneslksklehankensdvialwkskletaishqgameelkvsfskglgt 640
QY 327 YSSDRQIFCQ-----ACLRPEPDSENCLISRLMLWDNAKLYKGARKILHELI 374
DB 641 ecaefaelktqekmrlvdyheienlqngqdsr----- 674
QY 375 FSSFFMEYKYLFAFEFYKYLQKQKEYISDDHRSISITALSVMQFTVPTLARHLEE 434
DB 675 -aahakemealraklmkvik-ekensleairskldkaed-----qhlvem 717
QY 435 QNVISVITET-----LLEVLPEYLDNRNKNFQGVSQDKLGRVYAVICDLKYILISKPTI 489
DB 718 edtlmlqaeaeikvkelevlqakcneqtkvndftsq--lkateekildl----- 765
QY 490 WTERLRMQFLEGRFSLKLTTCMQGMEETRRQVGHIEVDPDWEAAIAI-----QMOLKNI 545
DB 766 --dalrkasseq-ksemkkl--rqgleaekqi-khleiekaeassknsntrelq---- 815
QY 546 LLMFQEWACADELLLVAYKECHKAVMRCSTFISSSKTVVOSCGHSLSTKSVRVSEDLV 605
DB 816 -----grelkltnlqe-----nlsevsqvketelekqilkekefaeaeav 857
QY 606 SIHLPLSRTLAGHLRLSLRLGAVSLHFEVSPEDFQVEVLVEYPLRCLVLVAQVVAEMWR 665
DB 858 svqsrsmqetvnlkhqkeefmlls-----dlekl----- 887
QY 666 RNGLSLISQVYYQD-----VKCREMYDKDI--IMLQIGASLMDPNKFLLLVLRQYEL 717
DB 888 renladmeakfrekdeeeqlikakekl-endiaieimknsdgndssqltk----- 935
QY 718 REAFNKTISTKDDLIKQNTLIEMLQVLIYIGERYVPGVGNVTKEEVTMRITHLIC 777
DB 936 ---mndeirlkerdy-----eelqltkltkanen--asflqksiedmtvk----- 974

QY 778 IEPMPHSAIAKNLPENNNETGLENVINKVATFKPGVSGHGVYELKDESLKDFNMYFYH 837
DB 975 aeqsggeaakkheekelerklsdlekmet-----shnqcqelkar-----y 1018
QY 838 YSKTQHSKAEHMOKRRKQENKDEALPPPPPEFCFAPSKVINLLNCDIMMYTLRTVFER 897
DB 1019 eratsetktkhee-----llgnlqkt 1039
QY 898 AIDTDSNLWTEGMLQMAFHIALGLLEEKQQLQKAPEEVEVTFDYHKASRLGSSAMNI-- 955
DB 1040 lldtedkl--kgaree-----nsgllqeleeirkaadka-----kaaqaedamqime 1085
QY 956 QMLLEKLGIPQLEGQKDMITWILQMFQVRLRKRKSLIVATTSSESINKDEIHTDKKE 1015
DB 1086 qmtkettetlasledtkqtnaklgneldtlkennlknveelnkskelltvengmeefrk 1145
QY 1016 KAERKRAEAAARLHRQKIMQAQMSALQKNFIETHKLMYDNTSEMPGKEDSI----MEEES 1070
DB 1146 eiellkqaadq-----sqqlsalqe---envkl-----aeelgrsrdevstshqleer 1192
QY 1071 TPAVSDYSRIALGPKRGPSTVTEKEVLTCLCOEEQEVKIENNAMVLSACV-QKSTALTQH 1129
DB 1193 ---svlnnqllemkkreskfikda-----deekasiqksisitsalltekdadelekl 1241
QY 1130 RGKPIELSGE 1139
DB 1242 rnevtvirge 1251
RESULT 15
AAM39097
ID AAM39097 standard; Protein; 2663 AA.
XX AAM39097;
DT 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 2242.
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW leukochemkinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-052317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AA158253.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:32:35 ; Search time 2649.68 Seconds
(without alignments)
25582.097 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattcgccagcagg.....aatttgattgggtttt 6308

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_eston:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	674.6	10.7	682	10	AI929033 au64c10.y
2	656.6	10.4	756	11	BG382624
3	637.4	10.1	641	10	AI361043 qy03f11.x
4	627.2	9.9	797	11	BG862813
5	573.2	9.1	782	11	BI086469 602849734
6	557	8.8	565	11	BF063405 7h89406.x
7	512.2	8.1	606	10	BES89438 195602.BA
8	479.8	7.6	624	11	BG219270 RST39023
9	477.8	7.6	481	10	AI192195
10	473.4	7.5	583	10	AW971391 EST383480.x
11	471.6	7.5	478	10	AA401319
12	454	7.2	522	10	AW291190 UI-H-B12-

13	449	7.1	522	11	BG382648
14	448	7.1	520	11	BG382624
15	410.8	6.5	486	10	BG50873 UI-M-BH3-
16	354.8	5.6	925	10	BES73578
17	352	5.6	374	10	AA400279
18	351.4	5.6	435	10	AA489271 UI-M-BH3-
19	330	5.2	394	11	BF774117 283680.MA
20	318	5.0	481	10	AI187306 qf28h08.x
21	307.2	4.9	397	10	AW311960 6005.MARC
22	301.8	4.8	504	10	AI980640 pat.pk002
23	286.2	4.2	327	10	BES54236 UI-M-AK1-
24	285.2	4.2	523	11	BG797647 ic16b02.x
25	263.4	4.2	300	11	Z17892 HSDHE1041.S
26	252.6	4.0	1079	11	BG292980 602389655
27	249	3.9	562	10	AI504731 vl13c11.x
28	243.6	3.9	372	11	BF542537 UI-R-C3-S
29	242.8	3.8	505	11	BG364916 dc93c08.y
30	241	3.8	325	10	AI615529 vl13c11.y
31	237	3.8	291	10	AV225341 AV225341
32	235.8	3.7	396	11	BES30958 RC3-GN007
33	232.8	3.7	256	10	AI921294 wo23a03.x
34	231.4	3.7	278	10	BE077143 RC5-BT060
35	230.6	3.7	323	11	H33916 EST110358.R
36	223.6	3.5	745	10	AW976158 EST388267
37	222	3.5	555	11	BE873236 601451771
38	217	3.4	685	10	AI693180 wd68e01.x
39	209	3.3	313	10	AV168252 AV168252
40	208.6	3.3	632	11	BG625558 pgnic.pk0
41	206.8	3.3	221	10	AA507138 nh42d07.S
42	204	3.2	1148	11	BF164318 601773090
43	201	3.2	620	10	AW702134 uq98d05.x
44	200.4	3.2	435	11	BES30879 RC3-GN007
45	200	3.2	711	10	AI646734 ud65b06.x

ALIGNMENTS

RESULT 1

AI929033 682 bp mRNA EST 23-AUG-1999
LOCUS au64c10.v1 Schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION IMAGE:2519538 5', similar to TR:070481 070481 UBIQUITIN-PROTEIN
LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.

ACCESSION AI929033 GI:5664997

VERSION AI929033.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 682)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Krillman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin

,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished (1997)

COMMENT Other_ESTs: au64c10.x1

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Seq primer: -40RP from Gibco

High quality sequence stop: 450.

Location/Qualifiers

1. 682

/organism="Homo sapiens"

```
/db_xref="taxon:9606"
/clone="IMAGE:2519538"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCGCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCGCCCCCCC-3'
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
1 others
BASE COUNT 178 a 155 c 166 g 182 t
ORIGIN

Query Match 10.7%; Score 674.6; DB 10; Length 682;
Best Local Similarity 99.3%; Pred. No. 3.9e-147;
Matches 67; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5261 aactccgctgaggaactgcataccaattctgcagaagagagacagtcactctgtag 5320
Db 1 AACTCCGCGCTGAGGAAGTGCATACCAATCTGCAGAAGGAGAGTACAGTGCACCTGTAG 60

Qy 5321 ctatctatcttaactacaattgttctctctcttcagggaatttgggatactgtga 5380
Db 61 CTATCTATCTTTTACCTACAAATTTGTCCTCTCTCTTCAGGAATATGGGATCTGTAG 120

Qy 5381 gccctgtctcagagtggtgtgcagatccctgccttacttaacttgcagcaaaaaa 5440
Db 121 GCCCTGTCTCAGAGGTGGTGTGCAGATCCCTGCTTACTTAACCTGTTGAAGCAAAAAA 180

Qy 5441 caccgtgtcaggtaccctgaagaaagaaatagtttgatagagcttctctgatgactatag 5500
Db 181 CACCGTGTCTCAGGTACCTCCTGAGAAAGAAATAGTTGTATAGAGCTTCTGTATGACTATAG 240

Qy 5501 ctgctctctgaatacagcttctcattcaggtgcccaggtctgcagatgatgagcgaaa 5560
Db 241 CTGCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCCGCTCTGCAGATGATGAGCGAAA 300

Qy 5561 gcattcctctctctctctctctggtggtctatactatgttctcagaacatttgcctca 5620
Db 301 GCATCTCTCTCTGCTCTCTCTGTTGGGCTATATCTATGTTCTCAGAACATTTGCTGCCA 360

Qy 5621 ggaattgtgaacggggaaggttgagcttgccttcttccacgcacttcaactgtggagc 5680
Db 361 GGAATTTGTGAACGGGGAAGGTTGGAGCTTGCAATTTTTCACGCACTTCACGTGGAGC 420

Qy 5681 cggagctgtcatttctctaaatacagagaatgccgagtggtcctggttgaaggttaagc 5740
Db 421 CGGAGCTGTGCATTTTCTCTAAATAATCAGAGAAATGCCAGTGGTCTGTGTTGAAGGTAAAGC 480

Qy 5741 cagaggctgtgcctatccagctccttaacttgatgaataatggagaaaacagacctggcct 5800
Db 481 CAGAGGCTGTGCCCTATCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCCCTGGCCT 540

Qy 5801 gaagaggggcaaccccttatttctcgtgagcggtatcggaagctccatttgctgtg 5860
Db 541 GAAGAGGGGCAACCCCTTCTTATCTCTGTGAGCGGTATCGGAAGCTCCATTTTGGTCTG 600

Qy 5861 gcaacaacactgcattatagaagagattgtctagggagcaagagactaatcagatgtatt 5920
Db 601 GCAACAACACTGCATTATAGAAGAGATTGCTTAGGAGCCACAGAGACTAATCAGATGTATT 660
```

```
Qy 5921 tggattcaactggcagttactg 5942
||||| I ||| |||||
Db 661 GGGATTCAACCTGCAGNTACTG 682

RESULT 2
BG534574 756 bp mRNA EST 03-APR-2001
602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
DEFINITION mRNA sequence.
ACCESSION BG534574
VERSION BG534574.1 GI:13526116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 756)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1465 row: a column: 07
High quality sequence stop: 751.
Location/Qualifiers
1..756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4663182"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgtcggcc); Site_2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 274 a 138 c 152 g 192 t
ORIGIN

Query Match 10.4%; Score 656.6; DB 11; Length 756;
Best Local Similarity 95.9%; Pred. No. 6.5e-143;
Matches 730; Conservative 0; Mismatches 19; Indels 12; Gaps 5;

Qy 2836 ggtatgaacttgcggaggcttttaacaagaccatattcacaagaccagattgatta 2895
Db 1 GGTATGAACCTGCCGAGGCTTTTAAACAAGACCATATCTACAAAAGACCAGGATTGATTA 60

Qy 2896 aacaataataacactaataagaagaatgctcaggctccatctatattgtgggtgagc 2955
Db 61 AACAATATATACACTAATAGAAGAAATGCTTCAGGTCCTCATCTATATTGTGGGTGAGC 120

Qy 2956 gtatgtacctgagtggaatgtgaccaaagaaggtcacaatgagagaaatcattc 3015
Db 121 GTATGTACTGGAGTGGGAATGTGACCAAGAAGAGGTGCACATGAGAGAAATCATTC 180

Qy 3016 acttgccttgcattgaaccatgccacacagtgccattgcccataaaatttacctgagaatg 3075
Db 181 ACTTGCTTTGTCATTGAACCCACACAGTGCCTAGTCCATTTGCCAAAATTTACCTCAGAATG 240
```

```
QY 3076 aaaaataagaaagtgcttagagaatgtcataaaacaaagtgccacatttaagaacaccag 3135
Db 241 AAAATAATGAACATGGCTTAGAGAATGTATATAACAAAGTGCCACATTTTAAGAAACCCAG 300
QY 3136 ggtatcaggccatggagtttatgaactaaagaatgaatcaactgaaagacttcaatatgt 3195
Db 301 GGTATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTGAAGACTTTCAATATGT 360
QY 3196 acttttatactactcaaaacccagcatagcaaggtgaaacatatgcagaaagaaagga 3255
Db 361 ACTTTTATCATATCTCAAAACCCAGCATAGCAAGCGTGAACATATGCGAGAGAAAGGA 420
QY 3256 gaaaacagaacaaagaatgaag-----cattgcccacacacacacactctcgaattct 3309
Db 421 GAAACAAAGAAACAAAGATGAAGTAAACATTTGCCGACCCACCCTCTCTGAATTCT 480
QY 3310 gccctgtttcagcaagtgatgaacttcaacttcaacttgatattcatgtatcatattctc 3368
Db 481 GCCCTGCTTTTCAGCAAGTGATTAACCTTCTCAACTGTGATATCATGATGTACATTCCTC 540
QY 3369 aggaccgtatttgagcgggcaatagacacagattctaaacttgagaccaaagggatgctc 3428
Db 541 AGACCGGTATTGAGCGGCAATAGACAGAGATTCTAACTTGTGACCAAGGGATGCTC 600
QY 3429 caaatggctttcatacttctgcatggtttactagaagaacacacagcttcaaaaa 3488
Db 601 CAAATGGTTTTCATATTTCTGCGATGGTTTACTTAGAGAGAAACACAG-TTCAAAAA 659
QY 3489 gctctgagaagaagtaacatttgactttatcatataaggcttcaagattgggaagtcca 3548
Db 660 GCTCCTGAAGAAGTAACATTTGACTTTTATCATATAAGGCTTCAAGATTGGGAAGTTCA 719
QY 3549 gccatgaatatcaaaatgcttttgaaacaaactcaaggaat 3589
Db 720 G-CATGAATATACAA---TGTTTTGGAATAATCAAAAGGATT 756

RESULT 3
LOCUS AI361043 641 bp mRNA EST 15-FEB-1999
DEFINITION qy03fil.x1 NCI-CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2010957 3'
similar to TR:P91133 P91133 SIMILAR TOS. CEREBISIAE
UBIQUITIN-PROTEIN LIGASE E3 COMPONENT SP:P19812. ;, mRNA sequence.
ACCESSION AI361043
VERSION AI361043.1 GI:4112664
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 641)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1083 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 453.
Location/Qualifiers
1. .641
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone_image="IMAGE:2010957"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGGCATATCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 174 a 155 c 135 g 175 t 2 others
ORIGIN

Query Match 10.1%; Score 637.4; DB 10; Length 641;
Best Local Similarity 99.5%; Pred. No. 2e-138;
Matches 638; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5457 cctagaaaaaagaatagtttgatagagcttctctgatgactatagctcctcctgaatcaa 5516
Db 641 CCTAGAAAAAGAAATAGTTTGATAGAGCTTCTCTGATGACTATAGCTGCTCTGAAATCCA 582
QY 5517 gcttctcattcaagtgccacaggtctgcagatgatgaggaagaagcattcctgctctgc 5576
Db 581 GCTTCTCATNTCAGGTGCCACGGTCTGCAGATGATGAGCAAGAGATCCTGTCTCTCTGCG 522
QY 5577 ctttctctg9gggctatactatcttctcagaacatttgctgcccaggaattgtgaacggg 5636
Db 521 CTTTCTCTGGGGCTATACTATGTTCTCAGAACATTTGCTGCCAGAGAAATTTGTGAACGGG 462
QY 5637 gaagaggttgagcttgctcatcttttcacgcacttcactgtgagccgagctgcatttcc 5696
Db 461 GAAGAGTTGGAGCTTGTCATTTTTCAGGCACCTTCACGTGAGCGCGAGTCTGCATTTTC 402
QY 5697 ctaaaaatcagagaatgcccagtgctgctggttgaaagtgtaaaagccagaggtgtgctcat 5756
Db 401 CTAAAAATCAGAGAAATGCCAGTGGTCTGTTGAAGGTAAAGCCAGAGGCTGTGCCCTAT 342
QY 5757 ccagctcccttacttgatgaataatgagaacacagaccctgcccctgagaggggcaacccc 5816
Db 341 CCAGCTCCTTACTTGGATGAATATGGAGAACAGACCCTGCCCTGAAGAGGGGCAACCCC 282
QY 5817 cttcatttatctgtagcggtatcggaagctccatttggctgccaacacactgcatt 5876
Db 281 CTTTATTATCTGTCAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACACACTGCATT 222
QY 5877 atagaagagattgctaggagcccaagactaatcaatcagatgttatttgattcaactggcag 5936
Db 221 ATAGAAGAGATTGCTAGGAGCCCAAGACTTAATCAGATGTTATTTGGATTCAACTGGCAG 162
QY 5937 ttactgtgactccaaactctgcctcaagaacaaatcaacaaagacacagtaataaggctg 5996
Db 161 TTACTGTGAGCTCCAACCTCTGCCTCAAGCAATACAAATACACACACTAGTAGTAAGGCTG 102
QY 5997 attcaaatattgaaaaactttctgagggctgggaagatttgagggtcttttgctcca 6056
Db 101 ATTCAAATATTGAAAAACTTTCTGAGGGCTGGGAAGTAGTTGGAGGCTCTTTTGCTCCA 42
QY 6057 tgtccaggttccattcacatcaataaaaaattttcttaattgga 6097
Db 41 TGTCAGGTTCACTTACATCAATAAAAAATATTTCTTAATGGA 1

RESULT 4
LOCUS BG862813
DEFINITION BG862813 797 bp mRNA EST 29-MAY-2001
602799074F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4934370 5',
rna sequence.
ACCESSION BG862813
```